GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 4986.97 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_1_900
Perfect score: 900
Sequence: 1 tgggtatcccgagccgctgg.....cttgcgcacctgcgcgaaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	900	100.0	84868	1	DQ013294	DQ013294 Polyangiu
c	2	334.4	37.2	6979	1	AB045630	AB045630 Pseudomon
c	3	334.4	37.2	10929	1	AE004886	AE004886 Pseudomon
c	4	327.6	36.4	1416	6	AR148436	AR148436 Sequence
c	5	327.6	36.4	1416	6	AR177918	AR177918 Sequence
c	6	327.6	36.4	1416	6	CS131758	CS131758 Sequence
c	7	327.6	36.4	4377	6	AR148449	AR148449 Sequence
c	8	327.6	36.4	4377	6	AR177931	AR177931 Sequence
c	9	327.6	36.4	4377	6	CS131783	CS131783 Sequence
c	10	310.2	34.5	1861	1	AF155934	AF155934 Alcaligen
	11	298.6	33.2	110000	1	AE016853_10	Continuation (11 o
c	12	297.6	33.1	4957	1	AY929300	AY929300 Pseudomon
c	13	296	32.9	300933	1	AE016791	AE016791 Pseudomon
c	14	283	31.4	110000	1	CP000076_60	Continuation (61 o
	15	282.6	31.4	110000	1	CP000075_09	Continuation (10 o
	16	281.4	31.3	110000	1	BX571965_27	Continuation (28 o
	17	281.4	31.3	110000	1	CP000010_18	Continuation (19 o
c	18	279.8	31.1	110000	1	BX571965_18	Continuation (19 o
	19	276.2	30.7	110000	1	CP000058_10	Continuation (11 o
c	20	274.8	30.5	10565	1	AE004621	AE004621 Pseudomon
c	21	267.8	29.8	972	6	AR627614	AR627614 Sequence
c	22	267.8	29.8	1053	6	AR624263	AR624263 Sequence
c	23	267.8	29.8	3517	1	AY388473	AY388473 Myxococcu
	24	267.8	29.8	11812	6	AR619646	AR619646 Sequence
	25	267.8	29.8	12849	6	AR619568	AR619568 Sequence
	26	265.8	29.5	300242	1	AE016790	AE016790 Pseudomon
c	27	264.6	29.4	1617	6	AR623983	AR623983 Sequence
	28	264.6	29.4	19053	6	AR653687	AR653687 Sequence
c	29	264.6	29.4	34552	6	AR619867	AR619867 Sequence
	30	264.4	29.4	13540	1	AE004661	AE004661 Pseudomon
c	31	263.6	29.3	110000	1	AP006840_04	Continuation (5 of
c	32	261.2	29.0	216050	1	AL646076	AL646076 Ralstonia
	33	260.8	29.0	110000	1	CR555306_17	Continuation (18 o
c	34	260	28.9	1362	6	AR623753	AR623753 Sequence
c	35	260	28.9	29559	6	AR619859	AR619859 Sequence
c	36	256.6	28.5	3304	1	AB091764	AB091764 Pseudomon
c	37	256.2	28.5	1341	1	AY337515	AY337515 Myxococcu
c	38	256.2	28.5	1344	6	AR626157	AR626157 Sequence
	39	256.2	28.5	6155	6	AR619342	AR619342 Sequence
	40	256.2	28.5	110000	1	CR555306_00	CR555306 Azoarcus
c	41	254.6	28.3	3699	6	AR622330	AR622330 Sequence
	42	254.6	28.3	19112	6	AR619786	AR619786 Sequence
	43	253.2	28.1	10526	1	AF325554	AF325554 Ralstonia
	44	252.6	28.1	7775	1	AB063252	AB063252 Burkholde
c	45	250.4	27.8	1107	1	AY337512	AY337512 Myxococcu

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRARRAABREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPFFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
CDS complement(1992. .3440)
/note="ORFu2; similar to Pseudomonas putida KT2440 PP4696; putative nitrogen regulation protein NR(I)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="AAY32957.1"
/db_xref="GI:63099934"
/translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
ELTETEIARRLGISRKALWERRQRLGIPRPRS"
CDS 3587. .5548
/note="ORFu3; ORFu3p; similar to Leptospira interrogans AAN50474"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAY32958.1"
/db_xref="GI:63099935"
/translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
VVAGFGATGPLRVMETLPEQYLFLTRVRP"
CDS complement(5702. .7231)
/note="ORFu4; ORFu4p; similar to Burkholderia fungorum hypothetical protein in ZP_00033665"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAY32959.1"
/db_xref="GI:63099936"
/translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPPFSANQRPASTLAQWRVGLEHP
GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
GRVITVHFHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
P"
CDS complement(7561. .8931)
/note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440 PP4696; putative nitrogen regulation protein NR(I)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="AAY32960.1"
/db_xref="GI:63099937"
/translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
PPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
IPRPRS"
CDS complement(8928. .11939)
/note="ORF1; ORF1p; similar to Pseudomonas putida KT2440 HisK (PP4695) in GenBank Accession Number AAN70268"
/codon_start=1
/transl_table=11
/product="putative sensory box histidine kinase"
/protein_id="AAY32961.1"
/db_xref="GI:63099938"
/translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
FTAFFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPARERERAAR
LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
CDS complement(12405. .13379)
/note="ORF2; ORF2p; similar to family of known or putative phosphotransferases, including macrolide 2-phosphotransferases; similar to Bacillus subtilis YcbJ, Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273 MphB and Staphylococcus aureus plasmid pSR1 MphBM"
/codon_start=1
/transl_table=11
/product="putative macrolide kinase"
/protein_id="AAY32962.1"
/db_xref="GI:63099939"
/translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
CDS 13368. .14663
/note="ORF3; ORF3p"
/codon_start=1
/transl_table=11
/product="putative Ser/Thr protein kinase"
/protein_id="AAY32963.1"
/db_xref="GI:63099940"
/translation="MPGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
gene 15731. .33766
/gene="dszA"
CDS 15731. .33766
/gene="dszA"
/note="PKS (polyketide synthase) subunit modules 1-4a"
/codon_start=1
/transl_table=11
/product="DszA"
/protein_id="AAY32964.1"
/db_xref="GI:63099941"
/translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
PRAHYSPDFNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT

```
                     LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                     LSPDGLCKTFDADANGYVPGDGVGVLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                     ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                     RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                     VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                     RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                     DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                     RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                     RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                     ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                     RAQFTFKKFLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                     LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                     PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                     PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                     EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE

  Query Match          100.0%;  Score 900;  DB 1;  Length 84868;
  Best Local Similarity  100.0%;  Pred. No. 5.8e-80;
  Matches  900;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGGGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7575 TGGGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGAC 7634

Qy         61 GGGCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7635 GGGCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGC 7694

Qy        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7695 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 7754

Qy        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7755 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 7814

Qy        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7815 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 7874

Qy        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7875 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 7934

Qy        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7935 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCG 7994

Qy        421 GCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7995 GCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCT 8054

Qy        481 CGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8055 CGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCC 8114

Qy        541 GGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8115 GGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCC 8174

Qy        601 CCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8175 CCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCG 8234

Qy        661 CCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8235 CCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGA 8294

Qy        721 ACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8295 ACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCT 8354

Qy        781 GCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8355 GCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCC 8414

Qy        841 CCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8415 CCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 8474
```

```
RESULT 2
AB045630/c
LOCUS       AB045630               6979 bp    DNA     linear   BCT 19-JUN-2001
DEFINITION  Pseudomonas aeruginosa genes for glutamate-tRNA synthetase, CbrA,
            CbrB, complete cds.
ACCESSION   AB045630
VERSION     AB045630.1  GI:12082248
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (sites)
  AUTHORS   Nishijyo,T., Haas,D. and Itoh,Y.
  TITLE     The CbrA-CbrB two-component regulatory system controls the
            utilization of multiple carbon and nitrogen sources in Pseudomonas
            aeruginosa
  JOURNAL   Mol. Microbiol. 40 (4), 917-931 (2001)
   PUBMED   11401699
REFERENCE   2  (bases 1 to 6979)
  AUTHORS   Itoh,Y. and Nishijyo,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Yoshifumi Itoh, National Food Research
            Institute, Applied Microbiology; Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8642, Japan (E-mail:yosifumi@nfri.affrc.go.jp,
            Tel:81-298-38-8075, Fax:81-298-38-7996)
FEATURES             Location/Qualifiers
     source          1. .6979
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            284. .1165
                     /gene="gltX"
     CDS             284. .1165
                     /gene="gltX"
                     /note="putative"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="glutamate-tRNA synthetase"
                     /protein_id="BAB20865.1"
                     /db_xref="GI:12082249"
                     /translation="MTSSYVGRFAPTPSGYLHFGSLVAAVASYLDARAVGGRWLVRME
                     DLDPPREVPGAQQAILETLERYGFEWDGAVERQSERFPAYAAVVEQLLRSGLAYACTC
                     SRKQLEGFAGIYPGFCRDAGHAREDAAIRLRVPELEYRFVDRVQGEVRQHLGREVGDF
                     VIQRRDGLYAYQLAVVLDDAWQGITDIVRGADLLDSTPRQLYLQELLGLSQPRYLHVP
                     LIVQPDGHKLGKSYRSPPLPAEQAAAPLTRALRALGQRPPAELAQASASEALAWGVAH
                     WDATRIPRCATLPEERL"
     gene            1395. .4346
                     /gene="cbrA"
     CDS             1395. .4346
                     /gene="cbrA"
                     /function="signal sensor and protein histidine kinase"
                     /note="KinA-NtrB family
                     histidine at 766, phosphrylation site
                     histidine kinase domain,491 to 983 amino acids
                     sensor membrane domain, 1 to 490 amino acids"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CbrA"
                     /protein_id="BAB20866.1"
                     /db_xref="GI:12082250"
                     /translation="MLTSFSLTQLILISVTYLSTLFGIAWITERGYVPRRLVRHPLVY
```

```
                     TLSLGVYASAWAFYGTVGLAYQYGYGFLAIYLGISGAFLLAPVLLYPILRVTRAYQLS
                     SLADLFAFRFRSTWAGALTTLFMLIGVLPMLALQIQAVTDSISILTRDPAQERVAFVF
                     CTLITLFAILFGARHIATRERHEGLVFAIAFESLVKLVMLGSIGLYALYGVFGGPEGL
                     EVWLLQNQTALTTLHTPLAEGPWRTLLLVFFAAAIVMPHMFHMIFTENLNPRALISAS
                     WGLPLFLLLMSLAVPPILWAGLRLGASTTPEYYTIGLGLAVDSPALALAAFIGGISAA
                     SGLTIVMTLALSGMVLNHLVLPLYQPPAQGNIYRWLKWTRRLLIVAIIMASYAFYLLL
                     GAEQDLSNLGIVSFVATLQFLPGALSVLYWPTANRRGFISGLIAGMLVWAITMLLPLM
                     GNVQGLYLPLFDVIYVLDDSNWHLAALSSLAANVLVFTLVSLFTEASDEEKGAAEACA
                     VDNVRRPQRRELFAGSPQEFASQLAKPLGAKTAQKEVEQALRDLHLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDIVETFLPYKTSEESYVTEDIHFIESRLEDYHSRLTGLA
                     AELDTLRRYHRQTLQDLPMGVCSLAKDQEVLMWNRAIEELTGVGAQKVVGSRLSALPE
                     PWKGLLERFIDAPDEHLHKQRLVYDGHTRWLNLHKAAIEEPLAPGNSGLVLLVEDLTE
                     TQLLEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGELTEISEQ
                     ILDQTKRVSRIVQSLMSFAHSGSHLQALEPVCLSEVAQEAIGLLSLNRRSVEVEFFNL
                     CDPAHWVEGDSQRLAQVLINLLSNARDASPPGGAIRVRSEASEHTVDLVVEDEGSGIP
                     KAIMDQLFEPFFTTKDPGKGTGLGLALVYSIVEEHYGQITIESPTDHQREGGTRFRVT
                     LPRHPGPTAEL"
     gene            4373. .5809
                     /gene="cbrB"
     CDS             4373. .5809
                     /gene="cbrB"
                     /function="response regulator"
                     /note="Aspertate at 52 amino acids, phosphorylation site
                     NtrC family"
                     /codon_start=1
                     /transl_table=11
                     /product="CbrB"
                     /protein_id="BAB20867.1"
                     /db_xref="GI:12082251"
                     /translation="MAHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERYTIP
                     SFDMVVSDLRLPGAPGTELIKLAEGIPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILKDRQENRSAAPVSANGGKAGGERGASPAVADGEIGIIGSCAPMQELY
                     SKIRKVAPTDSTVLIQGESGTGKELVARALHNLSKRAKAPLISVNCAAIPETLIESEL
                     FGHEKGAPTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRVGSV
                     QSQKVDVRLIAATHRDLKTLAKTGQFREDLYYRLHVISLKLPALRERGNDVMEIARAF
                     LARQCTRMGRAPLSFAHDAEQAIRHYPWPGNVRELENAIERAVILSESQEIHADLLGI
                     DIELDDLEDDFAADLGLGPAGAAASNHEPTEDLSLEDYFQHFVLEHQDHMTETELARK
                     LGISRKCLWERRQRLGIPRRKSGASADS"
ORIGIN

  Query Match           37.2%;  Score 334.4;  DB 1;  Length 6979;
  Best Local Similarity  62.2%;  Pred. No. 2.3e-24;
  Matches  577;  Conservative    0;  Mismatches  321;  Indels   30;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || ||  || ||  |||||||| |||||||||| |   |||||||||||||| |||   |||
Db       5775 GGAATGCCCAGGCGCTGGCGGCGTTCCCACAGGCATTTGCGGCTGATGCCCAGCTTGCGG 5716

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  | ||||||||||||| || ||||||  |  |||| ||| |
Db       5715 GCCAGTTCGGTCTCGGTCATGTGGTCCTGGTGCTCGAGGACGAAGTGCTGGAAGTAGTCT 5656

Qy        123 TCGAGCGA--------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATG 162
              ||  |||||                    |   |  ||| ||| | |    ||   |||| |
Db       5655 TCCAGCGAAAGGTCCTCGGTGGGCTCGTGGTTGCTGGCCGCGGCGCCCGCCGGGCCGAGG 5596

Qy        163 CGGAGGAGACGGGCGGAGGCGGCCGCG----------GCGGGTCGTCGAGCCCGAGGTGG 212
              | ||||     |||| || || || ||          ||  |  |||||  || ||  ||
Db       5595 CCGAGGTCGGCGGCGAAGTCGTCCTCGAGGTCGTCCAGCTCGATGTCGATACCCAGCAGG 5536

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
               |    | |||  |   || | ||  ||| | ||| || || |  | |||||| |||||
Db       5535 TCGGCGTGGATTTCCTGGCTCTCGCTGAGGATCACCGCCCGCTCGATGGCGTTTTCCAGT 5476

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || || |||||||||||||| ||          ||| |||      || ||    |  |
Db       5475 TCGCGCACGTTGCCCGGCCATGGGTAATGGCGGATCGCCTGTTCGGCGTCGTGGGCGAAG 5416

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
               |||||||||  ||||||||          | | ||   ||| |  |    ||||||  |
Db       5415 CTGAGCGGCGCGCGCCCCATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCC 5356

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
```

```
                | | ||||   |   ||||| ||||| |  ||||||   | |   | |||  |||| |||
Db       5355 ATCACGTCGTTGCCACGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGG 5296

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
              ||||  |||||          |||||      |||  |      || || ||||| ||
Db       5295 TAATAGAGGTCTTCGCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTG 5236

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || ||||||| ||||||||  ||  | | |   ||  ||| || ||||| || |  |||
Db       5235 GCGGCGATCAGGCGCACGTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCG 5176

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||    |||||| |  |||||||||||  ||||  ||| || |||||| |||||| ||||
Db       5175 CCTTCCTGCAATACGCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCG 5116

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              ||||| | ||| || ||||| || ||||| ||| |||||| || ||| |  |||   ||
Db       5115 TCGAGGAACAGGGTGCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCA 5056

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
              || | |   |||||    || |  |||||||||||   ||||||  | |  || || |||
Db       5055 CCGGTGAAGGCGCCTTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTCCGGAATC 4996

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              ||||| || | |||     | | ||||  ||      |||      |||  |||||  |||
Db       4995 GCCGCACAGTTCACCGAGATCAGCGGCGCCTTGGCGCGCTTCGACAGGTTGTGCAGGGCG 4936

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || || | |   |||||||| |||||   |||||||   | ||| | ||| |  ||| ||
Db       4935 CGGGCGACCAGTTCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACCGTGGAGTCGGTC 4876

Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||| |||||||| | ||    | |||
Db       4875 GGCGCGACCTTGCGAATCTTGCTGTACA 4848
```

```
RESULT 3
AE004886/c
LOCUS       AE004886               10929 bp    DNA     linear   BCT 19-FEB-2003
DEFINITION  Pseudomonas aeruginosa PAO1, section 447 of 529 of the complete
            genome.
ACCESSION   AE004886 AE004091
VERSION     AE004886.1  GI:9950982
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa PAO1
  ORGANISM  Pseudomonas aeruginosa PAO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 10929)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
   PUBMED   10984043
REFERENCE   2  (bases 1 to 10929)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 10929)
  AUTHORS   .
```

```
CONSRTM    Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)
TITLE      Direct Submission
JOURNAL    Submitted (04-FEB-2003) Department of Molecular Biology and
           Biochemistry, Simon Fraser University, 8888 University Dr.,
           Burnaby, British Columbia V5A 1S6, Canada
COMMENT    ----------------------------------------------------------------
           This represents the February 3, 2003 version of the continually
           updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,
           from PseudoCAP (see http://www.pseudomonas.com for latest updates
           and links to alternate annotations). PseudoCAP is coordinated by
           Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
           E.W. Hancock (University of British Columbia, Canada). We welcome
           submission through www.pseudomonas.com of any proposed changes.

           'Protein name confidence' is used to rate our confidence of the
           accuracy of the protein name.
           Class 1: Function experimentally demonstrated in P. aeruginosa.
           Class 2: Function of highly similar gene experimentally
                demonstrated in another organism (and gene context consistent
                in terms of pathways its involved in, if known).
           Class 3: Function proposed based on presence of conserved amino
                acid motif, structural feature or limited sequence similarity
                to an experimentally studied gene.
           Class 4: Homologs of previously reported genes of unknown function,
                or no similarity to any previously reported sequences.
           ----------------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .10929
                     /organism="Pseudomonas aeruginosa PAO1"
                     /mol_type="genomic DNA"
                     /strain="PAO1"
                     /db_xref="taxon:208964"
     gene            127. .573
                     /gene="dksA"
                     /locus_tag="PA4723"
     CDS             127. .573
                     /gene="dksA"
                     /locus_tag="PA4723"
                     /note="Protein name confidence: Class 1 (function
                     experimentally demonstrated in P. aeruginosa)"
                     /codon_start=1
                     /transl_table=11
                     /product="suppressor protein DksA"
                     /protein_id="AAG08109.1"
                     /db_xref="GI:9950983"
                     /translation="MSTKAKQQSSQQMTRGFEPYQETKGEEYMSERMRAHFTAILNKW
                     KQELMEEVDRTVHHMQDEAANPPDPADRASQEEEFSLELRARDRERKLIKKIDETLQL
                     IEDEEYGWCDSCGVEIGIRRLEARPTATLCIDCKTLAEIREKQLGS"
     gene            641. .1522
                     /locus_tag="PA4724"
                     /note="synonym: yadB"
     CDS             641. .1522
                     /locus_tag="PA4724"
                     /note="Protein name confidence: Class 3 (function proposed
                     based on presence of conserved amino acid motif,
                     structural feature or limited sequence similarity to an
                     experimentally studied gene)"
                     /codon_start=1
                     /transl_table=11
                     /product="probable aminoacyl-transfer RNA synthetase
                     (class I)"
                     /protein_id="AAG08110.1"
                     /db_xref="GI:9950984"
                     /translation="MTSSYVGRFAPTPSGYLHFGSLVAAVASYLDARAVGGRWLVRME
                     DLDPPREVPGAQQAILETLERYGFEWDGAVERQSERFPAYAAVVEQLLRSGLAYACTC
                     SRKQLEGFAGIYPGFCRDAGHAREDAAIRLRVPELEYRFVDRVQGEVRQHLGREVGDF
                     VIQRRDGLYAYQLAVVLDDAWQGITDIVRGADLLDSTPRQLYLQELLGLSQPRYLHVP
                     LIVQPDGHKLGKSYRSPPLPAEQAAAPLTRALRALGQRPPAELAQASASEALAWGVAH
                     WDATRIPRCATLPEERL"
     gene            1752. .4703
                     /gene="cbrA"
                     /locus_tag="PA4725"
     CDS             1752. .4703
                     /gene="cbrA"
                     /locus_tag="PA4725"
                     /note="Protein name confidence: Class 1 (function
                     experimentally demonstrated in P. aeruginosa)"
                     /codon_start=1
                     /transl_table=11
                     /product="two-component sensor CbrA"
                     /protein_id="AAG08111.1"
                     /db_xref="GI:9950985"
                     /translation="MLTSFSLTQLILISVTYLSTLFGIAWITERGYVPRRLVRHPLVY
                     TLSLGVYASAWAFYGTVGLAYQYGYGFLAIYLGISGAFLLAPVLLYPILRVTRAYQLS
                     SLADLFAFRFRSTWAGALTTLFMLIGVLPMLALQIQAVTDSISILTRDPAQERVAFVF
                     CTLITLFAILFGARHIATRERHEGLVFAIAFESLVKLVMLGSIGLYALYGVFGGPEGL
                     EVWLLQNQTALTTLHTPLAEGPWRTLLLVFFAAAIVMPHMFHMIFTENLNPRALISAS
                     WGLPLFLLLMSLAVPPILWAGLRLGASTTPEYYTIGLGLAVDSPALALAAFIGGISAA
                     SGLTIVMTLALSGMVLNHLVLPLYQPPAQGNIYRWLKWTRRLLIVAIIMASYAFYLLL
                     GAEQDLSNLGIVSFVATLQFLPGALSVLYWPTANRRGFISGLIAGMLVWAITMLLPLM
                     GNVQGLYLPLFDVIYVLDDSNWHLAALSSLAANVLVFTLVSLFTEASDEEKGAAEACA
                     VDNVRRPQRRELFAGSPQEFASQLAKPLGAKTAQKEVEQALRDLHLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDIVETFLPYKTSEESYVTEDIHFIESRLEDYHSRLTGLA
                     AELDTLRRYHRQTLQDLPMGVCSLAKDQEVLMWNRAIEELTGVGAQKVVGSRLSALPE
                     PWKGLLERFIDAPDEHLHKQRLVYDGHTRWLNLHKAAIEEPLAPGNSGLVLLVEDLTE
                     TQLLEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGELTEISEQ
                     ILDQTKRVSRIVQSLMSFAHSGSHLQALEPVCLSEVAQEAIGLLSLNRRSVEVEFFNL
                     CDPAHWVEGDSQRLAQVLINLLSNARDASPPGGAIRVRSEASEHTVDLVVEDEGSGIP
                     KAIMDQLFEPFFTTKDPGKGTGLGLALVYSIVEEHYGQITIESPTDHQREGGTRFRVT
                     LPRHPGPTAEL"
     gene            4730. .6166
                     /gene="cbrB"
                     /locus_tag="PA4726"
     CDS             4730. .6166
                     /gene="cbrB"
                     /locus_tag="PA4726"
                     /note="Protein name confidence: Class 1 (function
                     experimentally demonstrated in P. aeruginosa)"
                     /codon_start=1
                     /transl_table=11
                     /product="two-component response regulator CbrB"
                     /protein_id="AAG08112.1"
                     /db_xref="GI:9950986"
                     /translation="MAHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERYTIP
                     SFDMVVSDLRLPGAPGTELIKLAEGIPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILKDRQENRSAAPVSANGGKAGGERGASPAVADGEIGIIGSCAPMQELY
                     SKIRKVAPTDSTVLIQGESGTGKELVARALHNLSKRAKAPLISVNCAAIPETLIESEL
                     FGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRVGSV
                     QSQKVDVRLIAATHRDLKTLAKTGQFREDLYYRLHVISLKLPALRERGNDVMEIARAF
                     LARQCTRMGRAPLSFAHDAEQAIRHYPWPGNVRELENAIERAVILSESQEIHADLLGI
                     DIELDDLEDDFAADLGLGPAGAAASNHEPTEDLSLEDYFQHFVLEHQDHMTETELARK
                     LGISRKCLWERRQRLGIPRRKSGASADS"
     gene            7067. .8470
                     /gene="pcnB"
                     /locus_tag="PA4727"
     CDS             7067. .8470
                     /gene="pcnB"
                     /locus_tag="PA4727"
                     /note="Protein name confidence: Class 2 (high similarity
                     to functionally studied protein); Alt protein name:
                     polynucleotide adenylyltransferase, plasmid copy number
                     protein"
                     /codon_start=1
                     /transl_table=11
                     /product="poly(A) polymerase"
                     /protein_id="AAG08113.1"
                     /db_xref="GI:9950987"
                     /translation="MLKKLIQTIRSPLRRPRAVRTTPEVIGNNQHSLRRDQFSRNAVK
                     VVEILQRAGYQAYVVGGCVRDQMLGIAPKDFDVATSATPEQVRAEFRNARIIGRRFKL
                     VHVHFGRRIIEVATFRAHHPEGQDDGDSRASSNESGRILRDNVYGSLEDDAQRRDFTI
                     NALYFDVTSERLLDYANGVHDIRNRLIRLIGDPEQRYLEDPVRMLRAVRFAAKLDFEI
                     EKHSAAPIRRLAPLLREIPSARLFDEVLKLFLAGRAERTFELLVEYELFAPLFPASAK
                     ALQANPDYTGKLIRQALANTDARIRQGKPVTPAFLFAALLWPALPARVAQLQEKGMPA
                     IPAMQEAAHELISEQCQRIAIPKRFTLPIREIWDMQERLPRRQGKRADLLLENPRFRA
                     GYDFLLLRESAGEETEGLGQWWTDYQEVSDSERRNMIRDLVSQEDGSAPRKRRRGGNS
                     GRRRRGPRKEGSGGSGE"
```

```
     gene            8467. .8955
                     /gene="folK"
                     /locus_tag="PA4728"
     CDS             8467. .8955
                     /gene="folK"
                     /locus_tag="PA4728"
                     /note="Protein name confidence: Class 2 (high similarity
                     to functionally studied protein); Alt protein name:
                     6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase,
                     7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase"
                     /codon_start=1
                     /transl_table=11
                     /product="2-amino-4-hydroxy-6-
                     hydroxymethyldihydropteridine pyrophosphokinase"
                     /protein_id="AAG08114.1"
                     /db_xref="GI:9950988"
                     /translation="MIQRVYVALGSNLAEPREQIQAALDAFERLPETRLVAVSPLYIS
                     DPLGPADQPRFVNGVAALDTNLAPLDLLDALQAIELEQGRVRDLRWGPRTLDLDILLF
                     GEQLLDLPRLKVPHYHMQARAFVLYPLADLAPDLRLPDGRHLPELLAACPFEGIERLP

   Query Match             37.2%;  Score 334.4;  DB 1;  Length 10929;
   Best Local Similarity   62.2%;  Pred. No. 1.9e-24;
   Matches  577;  Conservative    0;  Mismatches  321;  Indels   30;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || |||||||| |||||||||| |   ||||||||||||| |||   |||
Db       6132 GGAATGCCCAGGCGCTGGCGGCGTTCCCACAGGCATTTGCGGCTGATGCCCAGCTTGCGG 6073

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  | |||||||||||| || ||||||  |  |||| | |||
Db       6072 GCCAGTTCGGTCTCGGTCATGTGGTCCTGGTGCTCGAGGACGAAGTGCTGGAAGTAGTCT 6013

Qy        123 TCGAGCGA--------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATG 162
              || |||||                    |   | ||| ||| | |    ||   ||| |
Db       6012 TCCAGCGAAAGGTCCTCGGTGGGCTCGTGGTTGCTGGCCGCGGCGCCCGCCGGGCCGAGG 5953

Qy        163 CGGAGGAGACGGGCGGAGGCGGCCGCG----------GCGGGTCGTCGAGCCCGAGGTGG 212
              | ||||     |||| || || || ||         ||  |  |||||  || ||  ||
Db       5952 CCGAGGTCGGCGGCGAAGTCGTCCTCGAGGTCGTCCAGCTCGATGTCGATACCCAGCAGG 5893

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
              |    | |||  |   || | ||  ||| | ||| || || |  | |||||||||||||
Db       5892 TCGGCGTGGATTTCCTGGCTCTCGCTGAGGATCACCGCCCGCTCGATGGCGTTTTCCAGT 5833

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || ||  |||||||||||||| ||         ||| |||      || ||    |  |
Db       5832 TCGCGCACGTTGCCCGGCCATGGGTAATGGCGGATCGCCTGTTCGGCGTCGTGGGCGAAG 5773

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
              |||||||||  ||||||||       | | ||   ||| |  |    ||||||  |
Db       5772 CTGAGCGGCGCGCGCCCCATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCC 5713

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
              | | ||||    |||||  ||||| |  ||||||    | |    | |||  |||| |||
Db       5712 ATCACGTCGTTGCCACGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGG 5653

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
              ||||  |||||    |||||  |||        |||  |      || || ||||| ||
Db       5652 TAATAGAGGTCTTCGCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTG 5593

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || |||||||| ||||||||  ||  | | |   ||  ||| || ||||| || |  |||
Db       5592 GCGGCGATCAGGCGCACGTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCG 5533

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||    ||||||  |  |||||||||||  ||||  |||  || ||||||  ||||||||
Db       5532 CCTTCCTGCAATACGCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCG 5473

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              |||||  | |||  || |||||  || ||||| |||  ||||||  || |||  | |||   ||
Db       5472 TCGAGGAACAGGGTGCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCA 5413

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
```

```
              || | |   |||||     || |  ||||||||||||   ||||||  |  |   ||  || |||
Db       5412 CCGGTGAAGGCGCCTTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTCCGGAATC 5353

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              |||||  || | |||      |  | ||||  ||      |||     |||  |||||  |||
Db       5352 GCCGCACAGTTCACCGAGATCAGCGGCGCCTTGGCGCGCTTCGACAGGTTGTGCAGGGCG 5293

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || || | |   ||||||||  |||||    |||||||    | |||  | |||  |  ||| ||
Db       5292 CGGGCGACCAGTTCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACCGTGGAGTCGGTC 5233

Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900
              |||||  |||||||| | ||    | |||
Db       5232 GGCGCGACCTTGCGAATCTTGCTGTACA 5205


RESULT 4
AR148436/c
LOCUS       AR148436               1416 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 3 from patent US 6225106.
ACCESSION   AR148436
VERSION     AR148436.1  GI:15112526
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1416)
  AUTHORS   Gerritse,G. and Quax,W.J.
  TITLE     Expression system for altered expression levels
  JOURNAL   Patent: US 6225106-A 3 01-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .1416
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             36.4%;  Score 327.6;  DB 6;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 1.7e-23;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || |||||| || || |||||||| |  |||||||||||||||||   ||
Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  | |||||||||||||| ||||||  |  |||| ||| ||
Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                  ||     |  ||  | |    |  ||  |    ||
Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  |  ||  |     |      | ||  |  |||||  || ||  |   |      ||
Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |    |||| ||  ||| | ||| ||||| |   | ||| ||||||||||| || ||||||
Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||      || |||   |  |      ||   | |
Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||   | |   | ||  || |  ||||||          ||||  ||
Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  ||||| |||||||||  ||    | |  ||  |||| ||| | || ||||||
Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
```

```
Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584

Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644

Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704

Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764

Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824

Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884

Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900

Db        482 CGGATCTTGCCGTAAA 467


RESULT 5
AR177918/c
LOCUS       AR177918               1416 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 3 from patent US 6313283.
ACCESSION   AR177918
VERSION     AR177918.1  GI:17920273
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1416)
  AUTHORS   Gerritse,G. and Quax,W.J.
  TITLE     Expression system for altered expression levels
  JOURNAL   Patent: US 6313283-A 3 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1416
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           36.4%;  Score 327.6;  DB 6;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 1.7e-23;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62

Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122

Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164

Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224

Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284

Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344

Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404

Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464

Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524

Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584

Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644

Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704

Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764

Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824

Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884

Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900

Db        482 CGGATCTTGCCGTAAA 467


RESULT 6
CS131758/c
LOCUS       CS131758               1416 bp    DNA     linear   PAT 02-AUG-2005
DEFINITION  Sequence 3 from Patent EP1555320.
ACCESSION   CS131758
VERSION     CS131758.1  GI:71794288
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Gerritse,G.
  TITLE     Expression system derived from the lipase regulation cascade of
            pseudomonas alcaligenes
  JOURNAL   Patent: EP 1555320-A 3 20-JUL-2005;
            Genencor International, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1416
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match           36.4%;  Score 327.6;  DB 6;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 1.7e-23;
```

```
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  |||||||||||||||||||    ||
Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| |||| |  |||||||||||||| |||||| |  |||| ||| ||
Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                ||      | ||  | |    |   || |     ||
Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
               ||  | ||  |    |      | ||  |  |||||  || ||  |  |        ||
Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
               |  |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||       || |||    |  |      ||   | |
Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
               | |||||   | |   | | ||    || |  |    ||||||         ||||  ||
Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||   | |  ||  |||| ||| | || |||||
Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||  ||          ||| || |  ||| || ||||| || ||  ||||||  |||||
Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   ||  ||| || ||||| || |  ||||||  |||||
Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  ||||||||||  ||||| ||| || ||||||||||||||  ||||||||| | |||
Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| ||| || || ||  | |||  ||||| | |   ||
Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    | | |||   ||    ||  |   ||  ||||| ||| ||||| | |  |
Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900
              || | ||  || | |
Db        482 CGGATCTTGCCGTAAA 467


RESULT 7
AR148449/c
LOCUS       AR148449               4377 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 28 from patent US 6225106.
ACCESSION   AR148449
VERSION     AR148449.1  GI:15112539
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4377)
  AUTHORS   Gerritse,G. and Quax,W.J.
  TITLE     Expression system for altered expression levels
  JOURNAL   Patent: US 6225106-A 28 01-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .4377
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             36.4%;  Score 327.6;  DB 6;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 1.2e-23;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  |||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| |||| |  |||||||||||||| |||||| |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                ||      | ||  | |    |   || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
               ||  | ||  |    |      | ||  |  |||||  || ||  |  |        ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
               |  |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||       || |||    |  |      ||   | |
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
               | |||||   | |   | | ||    || |  |    ||||||         ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||   | |  ||  |||| ||| | || |||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||  ||          ||| || |  ||| || ||||| || ||  ||||||  |||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   ||  ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  ||||||||||  ||||| ||| || ||||||||||||||  ||||||||| | |||
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| ||| || || ||  | |||  ||||| | |   ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    | | |||   ||    ||  |   ||  ||||| ||| ||||| | |  |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985
```

```
Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              ||||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909


RESULT 8
AR177931/c
LOCUS       AR177931               4377 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 28 from patent US 6313283.
ACCESSION   AR177931
VERSION     AR177931.1  GI:17920286
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4377)
  AUTHORS   Gerritse,G. and Quax,W.J.
  TITLE     Expression system for altered expression levels
  JOURNAL   Patent: US 6313283-A 28 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .4377
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           36.4%;  Score 327.6;  DB 6;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 1.2e-23;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| |||  |  |||||||||||||| |||||| |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                 ||     | ||  | |   |  || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |    |      | ||  |  |||||  || ||  |  |       ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||      || |||   |  |      ||   | |
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||  | |   | | ||   || |  |    ||||||          ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||| |||||  ||         ||| ||  | |  ||  |||| ||| | || |||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||   ||          ||| || |  ||| || ||||| || ||  |||||||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |  || ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  ||||||||||||  ||||| ||| || ||||||||||||||| |||||||| | |||
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  ||| || || ||  | |||     ||||| | |   ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | ||
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||   ||      ||  |    ||   ||||| ||| ||||| | | |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              ||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909


RESULT 9
CS131783/c
LOCUS       CS131783               4377 bp    DNA     linear   PAT 02-AUG-2005
DEFINITION  Sequence 28 from Patent EP1555320.
ACCESSION   CS131783
VERSION     CS131783.1  GI:71794301
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Gerritse,G.
  TITLE     Expression system derived from the lipase regulation cascade of
            pseudomonas alcaligenes
  JOURNAL   Patent: EP 1555320-A 28 20-JUL-2005;
            Genencor International, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .4377
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match           36.4%;  Score 327.6;  DB 6;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 1.2e-23;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| |||  |  |||||||||||||| |||||| |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                 ||     | ||  | |   |  || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |    |      | ||  |  |||||  || ||  |  |       ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525
```

```
Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
Db       1924 CGGATCTTGCCGTAAA 1909
```

```
RESULT 10
AF155934/c
LOCUS       AF155934               1861 bp    DNA     linear   BCT 13-JUN-1999
DEFINITION  Alcaligenes faecalis NifA gene, partial cds.
ACCESSION   AF155934
VERSION     AF155934.1  GI:5052045
KEYWORDS    .
SOURCE      Alcaligenes faecalis
  ORGANISM  Alcaligenes faecalis
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Alcaligenaceae; Alcaligenes.
REFERENCE   1  (bases 1 to 1861)
  AUTHORS   Yin,H., Lin,M. and Lu,W.
  TITLE     Study on the regulation of expression of nifA gene in Alcaligenes
            faecalis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1861)
  AUTHORS   Lin,M., Yin,H. and Lu,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Biotechnology, Institution for Application
            of Atomic Energy, Yuan Ming Yuan Westroad 2, Beijing 100094, P. R.
            China
FEATURES             Location/Qualifiers
     source          1. .1861
                     /organism="Alcaligenes faecalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:511"
     CDS             <1. .1288
                     /codon_start=2
                     /transl_table=11
                     /product="NifA"
                     /protein_id="AAD38416.1"
                     /db_xref="GI:5052046"
                     /translation="PGFDLIVSDLRLPGAPGTELIKLAEGTPVLIMTSYASLRSAVDS
                     MKMGAVDYIAKPFDHDEMLQTVARILQDHQQSAAALAQPSRGNGSAPSDAAQGDIGII
                     GHCAPMQDLFSKIRKVAPTDSNVLIQGESGTGKELVARALHNLSRRAKAPMISVNCAA
                     IPETLIESELFGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQ
                     EGEIRRVGSVQSQKVDVRLIAATHRDLKTLAKTGQFREDLYYRLHVIALKLPPLRERG
                     NDVLEIAKAFLARQGERMSSDEMRFSRDAEQAIRHYSWPGNVRELENAIERAAILSES
                     PEIDAELLGIDIELDELDDEFEEDSLGGIRGIAGGNEPTEDLSLEDYFQHFVLEHQDH
                     MTETELARKLGISRKCLWERRQRLGIPRRKSSSAAG"
ORIGIN
```

```
  Query Match             34.5%;  Score 310.2;  DB 1;  Length 1861;
  Best Local Similarity   60.5%;  Pred. No. 7.9e-22;
  Matches  560;  Conservative    0;  Mismatches  338;  Indels   27;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
Db       1257 GGAATGCCCAGGCGCTGGCGGCGTTCCCACAGGCATTTGCGGCTGATGCCGAGCTTGCGC 1198

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
Db       1197 GCCAGTTCGGTCTCGGTCATGTGGTCCTGATGCTCGAGCACGAAATGCTGGAAGTAGTCC 1138

Qy        123 TCGAGCGAGTCCGAAGGCGGCGCCCCGTCGC--GCAGCGATGCGGAGGAGACGGGCGGAG 180
Db       1137 TCCAGCGAGAGATCCTCGGTCGGCTCGTTGCCACCAGCGATACCGCGGATGCCGCCGAGT 1078

Qy        181 GCGGCCGCGGCG----------------------GGTCGTCGAGCCCGAGGTGGGCC 215
Db       1077 GAATCTTCTTCGAATTCATCGTCCAGCTCGTCGAGTTCGATATCGATGCCCAGCAGCTCG 1018

Qy        216 CTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCC 275
Db       1017 GCGTCGATCTCCGGGCTCTCGCTGAGGATCGCCGCGCGCTCGATGGCGTTTTCCAGTTCG 958

Qy        276 CGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCG 335
Db        957 CGCACGTTGCCGGGCCAGGAGTAGTGACGAATCGCCTGTTCGGCGTCGCGCGAGAAGCGC 898

Qy        336 AGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACC 395
Db        897 ATCTCGTCGCTGCTCATGCGTTCGCCCTGGCGCGCCAGGAACGCCTTGGCGATTTCGAGC 838

Qy        396 GCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAA 455
Db        837 ACGTCGTTGCCGCGCTCGCGCAGCGGCGGCAGCTTGAGCGCGATCACGTGCAGGCGGTAA 778

Qy        456 TACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCC 515
Db        777 TAAAGGTCTTCGCGGAACTGGCCGGTCTTGGCAAGGGTTTTCAGGTCGCGGTGGGTGGCG 718

Qy        516 GCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCGCACTTCGCCC 575
Db        717 GCGATCAGACGCACATCGACCTTCTGCGACTGCACCGAGCCGACGCGGCGGATCTCACCT 658

Qy        576 GTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCG 635
Db        657 TCCTGCAGTACGCGCAGCAGGCGCGCCTGGGCTTCCAGTGGCAGTTCGCCGATTTCGTCG 598

Qy        636 AGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCC 695
Db        597 AGGAACAGGGTGCCGCCATCGGCTGCCTCGACCAGGCCGGCGCGACCGGCGCTGGCCCCG 538

Qy        696 GCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCC 755
Db        537 GTGAAGGCCCCCTTCTCGTGGCCGAACAGCTCGGATTCGATCAAGGTTTCCGGAATGGCC 478

Qy        756 GCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGC 815
```

```
Db        477 GCGCAGTTCACCGAGATCATCGGCGCCTTGGCCCGTCGCGAGAGATTGTGCAGTGCCCGC 418

Qy        816 GCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGC 875
              |||| |   |||||||| ||||||  |||||||   | |||||   | |  ||| | ||
Db        417 GCCACCAGTTCCTTGCCGGTGCCCGACTCGCCCTGGATCAGCACGTTGGAATCGGTTGGG 358

Qy        876 GCCACCTTGCGCACCTGCGCGAACA 900
              || |||||||| | ||    |||||
Db        357 GCGACCTTGCGGATCTTGCTGAACA 333


RESULT 11
AE016853_10
WPCOMMENT
  Sequence split into 64 fragments  LOCUS AE016853 Accession AE016853
     Fragment Name      Begin      End
     AE016853_00            1   110000
     AE016853_01       100001   210000
     AE016853_02       200001   310000
     AE016853_03       300001   410000
     AE016853_04       400001   510000
     AE016853_05       500001   610000
     AE016853_06       600001   710000
     AE016853_07       700001   810000
     AE016853_08       800001   910000
     AE016853_09       900001  1010000
     AE016853_10      1000001  1110000
     AE016853_11      1100001  1210000
     AE016853_12      1200001  1310000
     AE016853_13      1300001  1410000
     AE016853_14      1400001  1510000
     AE016853_15      1500001  1610000
     AE016853_16      1600001  1710000
     AE016853_17      1700001  1810000
     AE016853_18      1800001  1910000
     AE016853_19      1900001  2010000
     AE016853_20      2000001  2110000
     AE016853_21      2100001  2210000
     AE016853_22      2200001  2310000
     AE016853_23      2300001  2410000
     AE016853_24      2400001  2510000
     AE016853_25      2500001  2610000
     AE016853_26      2600001  2710000
     AE016853_27      2700001  2810000
     AE016853_28      2800001  2910000
     AE016853_29      2900001  3010000
     AE016853_30      3000001  3110000
     AE016853_31      3100001  3210000
     AE016853_32      3200001  3310000
     AE016853_33      3300001  3410000
     AE016853_34      3400001  3510000
     AE016853_35      3500001  3610000
     AE016853_36      3600001  3710000
     AE016853_37      3700001  3810000
     AE016853_38      3800001  3910000
     AE016853_39      3900001  4010000
     AE016853_40      4000001  4110000
     AE016853_41      4100001  4210000
     AE016853_42      4200001  4310000
     AE016853_43      4300001  4410000
     AE016853_44      4400001  4510000
     AE016853_45      4500001  4610000
     AE016853_46      4600001  4710000
     AE016853_47      4700001  4810000
     AE016853_48      4800001  4910000
     AE016853_49      4900001  5010000
     AE016853_50      5000001  5110000
     AE016853_51      5100001  5210000
     AE016853_52      5200001  5310000
     AE016853_53      5300001  5410000
     AE016853_54      5400001  5510000
     AE016853_55      5500001  5610000
     AE016853_56      5600001  5710000
     AE016853_57      5700001  5810000
     AE016853_58      5800001  5910000
     AE016853_59      5900001  6010000
     AE016853_60      6000001  6110000
     AE016853_61      6100001  6210000
     AE016853_62      6200001  6310000
     AE016853_63      6300001  6397126
Continuation (11 of 64) of AE016853 from base 1000001 (AE016853 Pseudomonas syringae pv.

  Query Match           33.2%;  Score 298.6;  DB 1;  Length 110000;
  Best Local Similarity 59.3%;  Pred. No. 2.9e-21;
  Matches  545;  Conservative    0;  Mismatches  344;  Indels   30;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || |||||| || |||||||||| |  |||||| |||| ||| ||    ||
Db      47181 GGAATGCCCAGACGCTGACGGCGTTCCCACAGGCACTTGCGACTGACGCCCAGTTTGCGC 47240

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || | |||||| || ||||  | |||||| |||||||| || ||| |  | || ||| |
Db      47241 GCCAGCTCGGTTTCGGTCATGTGGTCCTGATGCTCCAGAACAAAGTGCTGAAAGTAGTCT 47300

Qy        123 TCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAGGC 182
              || |||||           | || | ||| ||    |||   ||| ||| | ||| | ||
Db      47301 TCAAGCGACAGGTCTTCGGCCGGCTCGTTGCCTGTCGAGCTGGAAGAGCCTGGCAGCGGT 47360

Qy        183 GGC----------------------------CGCGGCGGGTCGTCGAGCCCGAGGTGG 212
              | ||                          | |   |  |||||  || ||  ||
Db      47361 GCCATGCCCATGTACTCATCGTCGTCCAGACCGTCCAGTTCGATGTCGATACCCAGCAGG 47420

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
              |   |||  |  |  | ||  || | ||| | ||| |  |||||||| |||||
Db      47421 TCGGCGGTGATTTCCGGGTTCTCGCAAAGGATCACTGAGCGTTCCACGGCGTTTTCCAGT 47480

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || || |||||||||||||||           |||||||     || || |  |  | ||
Db      47481 TCGCGCACGTTGCCCGGCCAGCTGTAATGACGGATGGCCTGTTCGGCGTCGGGGGCAAAC 47540

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
              |||     | | || | |       || | |||   || | ||    ||| ||   |
Db      47541 TTGAGGTCGTCGCGACCGACCTTGGCGCTCTGGCGCACGAGAAACGCACGGGCAATTTCC 47600

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
              |   ||| | |   ||||| ||||| | |||||||   |  | |  ||  |||| |||
Db      47601 AGTATGTCGCTGCCACGCTCACGCAGGGCCGGCAGCTTCAGGGCTATGACGTGCAGGCGG 47660

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
              |  |||||||    || ||    || |       ||| || |  ||| |||||||| |||
Db      47661 TAGAACAGGTCTTCACGAAACTCGCCATTTTTGGCCAGCGTCTTGAGGTCCCGGTGGGTC 47720

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || |||||||| ||||| ||  ||  | | |   ||  ||| || || || || | |||
Db      47721 GCGGCGATCAGGCGCACATCGACCTTCTGCGATTGCACCGAGCCGACGCGTCGAATTTCA 47780

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||   |||||| |  ||||||||||  ||||  | | || ||||||  || ||||  ||
Db      47781 CCTTCCTGCAACACGCGCAGCAGGCGGGCCTGAGCTTCCAGCGGCAGTTCACCGATTTCA 47840

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              ||||| | |||||| ||||| || || || ||  ||| || || ||  | |||   ||
Db      47841 TCGAGAAACAGCGTGCCGCCGTCGGCTGCTTCAACCAGGCCTGCACGGCCTGCGCTGGCA 47900

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
              || | |  ||||||    ||||  |||||||  |||   |||||| |||  || |||||
Db      47901 CCGGTGAACGCGCCTTTTTCGTGACCGAACAATTCGGATTCGATCAGCGATTCCGGGATG 47960

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              |||||||| |  |||   | | ||||  ||      ||  |   |||   ||||| ||||
Db      47961 GCCGCGCAGTTGACGGAAATCATCGGCGCCTTGGCGCGGCGCGACAGGTTGTGCAGCGCG 48020

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || || | |  || || || |||||| |||||||   |||| |   | | ||| ||
Db      48021 CGGGCAACCAGTTCTTTACCGGTGCCCGACTCGCCCTGAACCAGAACGTTGGAGTCGGTC 48080
```

```
Qy        873 GGCGCCACCTTGCGCACCT 891
              || || || ||||| | ||
Db      48081 GGTGCGACTTTGCGAATCT 48099


RESULT 12
AY929300/c
LOCUS       AY929300               4957 bp    DNA     linear   BCT 12-JUL-2005
DEFINITION  Pseudomonas putida MhaS (mhaS) and MhaR (mhaR) genes, complete cds.
ACCESSION   AY929300
VERSION     AY929300.1  GI:62860911
KEYWORDS    .
SOURCE      Pseudomonas putida
  ORGANISM  Pseudomonas putida
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 4957)
  AUTHORS   Arias-Barrau,E., Sandoval,A., Naharro,G., Olivera,E.R. and
            Luengo,J.M.
  TITLE     A two-component hydroxylase involved in the assimilation of
            3-hydroxyphenyl acetate in Pseudomonas putida
  JOURNAL   J. Biol. Chem. 280 (28), 26435-26447 (2005)
   PUBMED   15866873
REFERENCE   2  (bases 1 to 4957)
  AUTHORS   Arias-Barrau,E., Luengo,J.M. and Olivera,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2005) Bioquimica Y Biologia Molecular,
            Universidad De Leon, Campus De Vegazana S/N, Leon 24007, Spain
FEATURES             Location/Qualifiers
     source          1. .4957
                     /organism="Pseudomonas putida"
                     /mol_type="genomic DNA"
                     /strain="U"
                     /db_xref="taxon:303"
     gene            421. .3396
                     /gene="mhaS"
     CDS             421. .3396
                     /gene="mhaS"
                     /note="two-component signal sensor"
                     /codon_start=1
                     /transl_table=11
                     /product="MhaS"
                     /protein_id="AAY16575.1"
                     /db_xref="GI:62860912"
                     /translation="MPMSFSLTQMILISAGYLMVLFGVAWISERGLIPRSIIRHPLTY
                     TLSLGVYASAWAFYGSVGLAYQYGYGFLACYLGVSGAFLLAPVLLYPILKITRTYQLS
                     SLADLLAFRFRSTWAGALTTIIMLIGVLPLLALQIQAVADSISILTGEPVKARVAFAF
                     CTLIILFTIFFGSRHIATREKHEGLVFAIAFESVIKLLALGGIGLYALYGVFGGPHEL
                     EVWLLQNQTALAALHTPLQEGPWRTLLLVFFASAIVMPHMYHMAFTENLNPRSLVSAS
                     WGLPLFLLLMSLAVPLVLWAGLRLGATTNPEYFTLGLGIAANNEALALLAYVGGLSAA
                     SGLIIVTTLALSGMALNHLVLPLYQPPAEGNIYRWLKWTRRALIVAIITAGFMFYLTQ
                     NNHQSLANLGIVAFVATLQFLPGVLSVLYWPTANRRGFIAGPLAGTLVWMVTMLLPLL
                     GNLQGPYIPLLDMIYVLDGTSWRMAAIASLAANVLLFTLISLFTNASTEEVSAAEACA
                     VDNVRRPQRRELHAASPQEFATQLAKPLGAKAAQKEVEQALRDLYLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDMVETFLPYKSGNENYVTEDIHFIESRLEDYHSRLTGLA
                     AELDALRRYHRQTLQELPMGVCSLAKDQEILMWNKAMEELTGIAAKHVVGSRLLTIDE
                     PWRGLLQGFINLPDEHLHKQRLALDGQPRWLNLHKPAIDEPLAPGNSGLVLLVEDLTE
                     TQALEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGEIIELSSQ
                     ILDQTKRVSRIVQSLMSFAHAGGSHQNSEEPVCLAEVAQDAIGLLALNRRNFEVQFFN
                     LCNPDHWAEGDPQRLAQVLINLLSNARDASPPGSAVRVRSEVSEHTVDLIVEDEGSGI
                     PKNIMDRLFEPFFTTKDPGEGTGLGLALVYSIVEEHYGQITIDSPADIERQRGTRIRV
                     TLPRHVVATSPEIRDRREN"
     gene            3401. .4843
                     /gene="mhaR"
     CDS             3401. .4843
                     /gene="mhaR"
                     /note="two-component response regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="MhaR"
                     /protein_id="AAY16576.1"
                     /db_xref="GI:62860913"
                     /translation="MPHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERFSIA
                     TFDLIVSDLRLPGAPGTELIKLGQGTPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILRDRQNAPAAASAAEPRATNGKAASADKGSAAAANGEIGIIGSCPPMQ
                     DMYSKIRKVAPTDSNVLIQGESGTGKELVARALHNLSRRAKAPMISVNCAAIPETLIE
                     SELFGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRV
                     GSVQSQKVDVRLIAATHRDLKNLAKAGQFREDLYYRLHVIALKLPALRERGSDVNEIA
                     SAFLARQSARIGRDDLHFSAEAEQAIRHYSWPGNVRELENAVERAVILSESAEISAEL
                     LGIDIELSDLEEDEILNNALVVASAASASHEPTEDLSLEDYFQHFVLEHQDHMTETEL
                     ARKLGVSRKCLWERRQRLGIPRRKSNATSE"
ORIGIN

  Query Match           33.1%;  Score 297.6;  DB 1;  Length 4957;
  Best Local Similarity   59.7%;  Pred. No. 9.9e-21;
  Matches  554;  Conservative    0;  Mismatches  344;  Indels   30;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || || || || |||||||||| |  |||||||||| ||| |||   ||
Db       4812 GGAATACCCAGGCGTTGACGGCGTTCCCACAGGCACTTGCGGCTGACGCCTAGCTTGCGT 4753

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  ||||| || ||||  | |||||||||||| |||||||||  |  |||| ||| |
Db       4752 GCCAGTTCGGTTTCGGTCATGTGGTCCTGGTGCTCGAGCACGAAGTGCTGGAAGTAGTCT 4693

Qy        123 TCGAGCGA---GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGA 179
              || |||||   |||    | |||| |   |  |   |  |  ||  ||  |||  | |
Db       4692 TCCAGCGAAAGGTCTTCGGTCGGCTCGTGGCTGGCACTGGCAGCACTGGCCACGACCAGC 4633

Qy        180 GGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCC------- 232
              |         |   |||||   | | |||| |  |  ||||||  || ||||
Db       4632 GCGTTGTTGAGGATCTCGTCCTCCTCCAGGTCGCTCAGCTCGATGTCGATGCCCAGCAGC 4573

Qy        233 --------------------CCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
                                   | ||  ||| | ||| |||||  |  |||||||| ||||||
Db       4572 TCGGCAGAAATTTCTGCGCTCTCGCTGAGGATCACCGCGCGCTCCACGGCGTTTTCCAGC 4513

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || || ||||| ||||||||          ||||||     || || |||||  |
Db       4512 TCACGTACGTTACCCGGCCAGCTGTAATGACGAATGGCCTGCTCGGCTTCGGCCGAGAAA 4453

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
                ||     |  | || ||   |    || | ||    || |  |    ||||||
Db       4452 TGCAGGTCGTCGCGGCCGATGCGCGCACTCTGGCGGGCGAGGAAGGCGCTGGCGATCTCG 4393

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
                |||| | |   ||||| ||||| |     |    | | ||| |||| |||  |||
Db       4392 TTGACGTCGCTGCCGCGCTCGCGCAGGGCAGGCAGTTTCAGGGCGATCACGTGCAGGCGG 4333

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
              | || |||||    |||||   ||  |    |||    |  || || ||||| |||
Db       4332 TAGTAAAGGTCCTCGCGGAACTGGCCGGCCTTGGCCAGGTTCTTCAGGTCGCGGTGGGTC 4273

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || |||||||| |||||  ||  ||    | |   ||  ||| || |||||  || |  |||
Db       4272 GCGGCGATCAGGCGCACATCGACCTTTTGCGATTGCACCGAACCTACCCGGCGAATCTCG 4213

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||   |||||  |  |||||||||||  ||||  | |  | |||||| |||||||  |||
Db       4212 CCTTCCTGCAGCACGCGCAGCAGGCGGGCCTGGGCTTCGAGCGGCAGTTCGCCGATTTCG 4153

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              ||||| | |||||| ||||| || || || ||  ||| || || ||| | |||   |||
Db       4152 TCGAGGAACAGCGTGCCGCCGTCGGCGGCTTCTACCAGGCCGGCACGCCCGGCGCTGGCG 4093

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
              |||| |  |||||    ||||  |||||||| ||   ||||||| | |  ||||||||
Db       4092 CCCGTGAAGGCGCCTTTTTCGTGACCGAACAGTTCTGACTCGATCAGTGTCTCGGGGATG 4033

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              || ||||| | |||     | | ||| ||    |||  |  |||  |||||  |||
Db       4032 GCGGCGCAGTTCACCGAGATCATCGGTGCCTTGGCCCGGCGCGACAGGTTGTGCAGGGCG 3973

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || |||| |  ||| || || || |||  |||||||   | |||||   | |  ||| |
Db       3972 CGGGCCACCAGCTCTTTACCGGTACCCGACTCGCCCTGGATCAGCACGTTGGAGTCGGTA 3913
```

```
Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||||||||||||| | ||    | |||
Db       3912 GGCGCCACCTTGCGGATCTTGCTGTACA 3885


RESULT 13
AE016791/c
LOCUS       AE016791              300933 bp    DNA     linear   BCT 12-DEC-2002
DEFINITION  Pseudomonas putida KT2440 section 18 of 21 of the complete genome.
ACCESSION   AE016791 AE015451
VERSION     AE016791.1  GI:26557036
KEYWORDS    .
SOURCE      Pseudomonas putida KT2440
  ORGANISM  Pseudomonas putida KT2440
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 300933)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Complete genome sequence and comparative analysis of the
            metabolically versatile Pseudomonas putida KT2440
  JOURNAL   Environ. Microbiol. 4 (12), 799-808 (2002)
REFERENCE   2  (bases 1 to 300933)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .300933
                     /organism="Pseudomonas putida KT2440"
                     /mol_type="genomic DNA"
                     /strain="KT2440"
                     /db_xref="taxon:160488"
     gene            153. .740
                     /locus_tag="PP4517"
     CDS             153. .740
                     /locus_tag="PP4517"
                     /note="identified by match to PFAM protein family HMM
                     PF03695"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAN70091.1"
                     /db_xref="GI:24986250"
                     /translation="MSFAEQLTRLQAFLDADELHEEALDYVAAHGYLTALSINAEDVP
                     EREWIDALFAEEPHYASEAQRTEIEATLVALKAHIARQLASDEEFDLPCDLDLTDEPD
                     DSDLRGWCIGFMEGVFLREEAWFENAEEEVSEMLLPIMVGSGLFDEQPEFADIASNAN
                     LQDDMIVQIPEALSALFLLLHAPDEKPALLKPRHH"
     gene            748. .1113
                     /locus_tag="PP4518"
     CDS             748. .1113
                     /locus_tag="PP4518"
                     /note="identified by match to PFAM protein family HMM
                     PF04304"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAN70092.1"
                     /db_xref="GI:24986251"
                     /translation="MRYLLLATGWLSVALGVLGIFLPVLPTTPFLLLAAACFARSSPR
                     FHHWLVHHPKLGPWIRDYLSGEGIPLKGKVYAIGLMWASIGLSCYLVPLFWARTFMLT
                     SAALVTLYILSLKTLHKPK"
     gene            1425. .2783
                     /gene="tolC"
                     /locus_tag="PP4519"
     CDS             1425. .2783
                     /gene="tolC"
                     /locus_tag="PP4519"
                     /note="similar to GB:X12387, SP:P05184, SP:P08684,
                     SP:P20815, PID:181354, PID:181372, PID:181374, PID:219963,
                     PID:35911, and PID:510086; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="agglutination protein"
                     /protein_id="AAN70093.1"
                     /db_xref="GI:24986253"
                     /translation="MRVLNPITSALLLALASANVQAMSITEAVQSAVDYHPQVSSNRN
                     SKLSADEDVKFARGGYYPSVDLVAGYGRQRSDNATTRAEGNHNKETLNYTQSELRLRQ
                     MIFDGFNTSNEVGRTEAVSTSRAYYTQAVAQDVALRAVEVYLEVLKRRELVTLAKNNL
                     QAHLRVNDQIGLRNERGVGSTADLDQSRARRALAENNLDTAEVDLADAEANFFSVIGR
                     APDELESPQSIKGEVPGTLEEARDGMRQNNPYIKSAQADVNAAEKQYEVGKSTFYPRF
                     DAILATGANNNTGGEKGHNNNDWQAGVEMNYNLFRGGSDKARLQSDAHKINQALDIRN
                     NALRELTENLSLAWNAMNNASKQLPTAREYAETTKRVRAAYQDQFGLGQRTLLDVLDS
                     ENELYNADRRYTEVRYTEEFSRYRVLATMGELLSKQHISLPPEALATTEVRTEARLPE
                     MR"
     gene            complement(2846. .3445)
                     /locus_tag="PP4520"
     CDS             complement(2846. .3445)
                     /locus_tag="PP4520"
                     /note="identified by match to PFAM protein family HMM
                     PF03942"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAN70094.1"
                     /db_xref="GI:24986254"
                     /translation="MPRPRCERCQRPLDHCLCPLIPALDSRTRVILLQHPSETAHALN
                     TARLAALGLNNAELRIGEVFDDLHDLLATPGYHPVLLFPGDDAQVLTAYGEAGDKPLL
                     LIVPDGTWRKARKLLYLNPLLGGLPRVTLGQVAASRYRLRKAPEAGALSTIEAVVGAL
                     NTLEQPACFDALLAPFETLIEGQIRAMGVETFQRNHGRE"
     gene            complement(3510. .5075)
                     /locus_tag="PP4521"
     CDS             complement(3510. .5075)
                     /locus_tag="PP4521"
                     /note="similar to GP:9947522; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="aerotaxis receptor, putative"
                     /protein_id="AAN70095.1"
                     /db_xref="GI:24986255"
                     /translation="MRMNLPVTEHERTFPGDQRLISTTDLDSRITYCNDAFVAISGFT
                     YDELVGQPHNLVRHPDMPPAVFGHMWETIKQGKPWMGIVKNRAKNGDYYWVSAYVTAI
                     YEQGRISGYESVRSVPTREQIRRAEALYARLRDGRSPVPWAARLGHGLGHGWPLIGAG
                     LLSAAGYLWLPPYAALAVLMASLLLAWYWVEHRQNQAIRRTLAEHPKAFTSPLVALTY
                     SDNPGLKGQLDLAIISEEARLQTALTRLVDAGVGVKSRAAQSSDLSDAQAQMLDRQRS
                     ETDQSATAIAQMAATIQQVTHNVQSTAHAASDADQLAQQGSELALKSLKDMGSMSDAV
                     NDIGQAVNALAEQTQSIGSVVDVITSIAEQTNLLALNAAIEAARAGEQGRGFAVVADE
                     VRSLAQRTRASTEEIHHIIASLRAGAERAVSTASRGEQISRDSVHSVEAVQAALSGIA
                     VAVSRITGMSQQMATASEQQSHVAEDINQQIVRIAQLCDQSAGQARQGAEISQDLERM
                     AEYLHSLAERFNR"
     gene            complement(5214. .6107)
                     /locus_tag="PP4522"
     CDS             complement(5214. .6107)
                     /locus_tag="PP4522"
                     /note="similar to GP:9947370; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, LysR family"
                     /protein_id="AAN70096.1"
```

```
                     /db_xref="GI:24986256"
                     /translation="MASTLPDLKLLRIFISVVRHQGFANAQRELNLSTSAISTYMSQL
                     EGALGIVLCHRGRGGFSLTSKGELFHQETLRLLAELDGFEQYAAALKGELRGTLNLGV
                     IDSTVGDRALPLAEAIGAYSQEHPAVHLHLSVSSPYELQLGVQDNRLDLAIGAFSSRM
                     SGLLYQPLYREQHWLYCSSRHPLYAERRIPEQVVTQQRMVGRGYWSQAELARHGFKHS
                     AATVESMEAQLILILSGAYIGYLPEHYAQAWVDKGDLRVLSPSTFGYQAPFSLIIRRG
                     RSREPLIQTFRDLLKSQLNVG"
     gene            6320. .7282
                     /locus_tag="PP4523"
     CDS             6320. .7282
                     /locus_tag="PP4523"
                     /note="similar to SP:P37819; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="agmatinase, putative"
                     /protein_id="AAN70097.1"
                     /db_xref="GI:24986257"
                     /translation="MRPTVDKTLHQPLGGNEMPRFGGIATMLRLPHLQSAKGLDAAFI
                     GVPLDIGTSLRSGTRFGPRQIRAESVMIRPYNMATGAAPFDSLSVADIGDVAINTFNL
                     LDAVRIIEEAYDEIVEHNVIPMTLGGDHTITLPILRALHKKHGKIGLVHIDAHADVND
                     HMFGEKIAHGTTFRRAVEEGLLDCDRVVQIGLRAQGYTADDFNWSRRQGFRVVQAEEC
                     WHKSLEPLMAEVREKVGGGPVYLSFDIDGIDPAWAPGTGTPEIGGLTTIQAMEIIRGC
                     HGLDLIGCDLVEVSPPYDTTGNTSLLGANLLFEMLCVLPGVVRR"
     gene            7415. .8794
                     /locus_tag="PP4524"
     CDS             7415. .8794
                     /locus_tag="PP4524"
                     /note="identified by match to PFAM protein family HMM
                     PF04279"
                     /codon_start=1
                     /transl_table=11
                     /product="sodium-solute symporter, putative"
                     /protein_id="AAN70098.1"
                     /db_xref="GI:24986258"
                     /translation="MALDIIVVMIYTAGMLGLGWYGMRRAKTHEDYLVAGRNLGPVLY
                     MGTMATTVLGGASTVGTVRLGYVHGISGFWLCAALGLGIIAINLFLAKPLLRLRIFTV
                     TQVLEQRYNPTARQASAVIMLAYALMIGVTSTLAMATVLQVLLDLPFWASLMLGGGVV
                     VLYSTIGGMWSLTLTDIVQFVIKTVGLMFILLPVCLYKAGGWDTLVAKLPAASFQLTT
                     IGWDTIITYFLIYFFGILIGQDIWQRVFTARDEKVCQRAGTAAGVYCVLYGLACAAIG
                     MAAHVLMPDLANPNNAFAEMIKTTLPDGIRGLLMAAALAAMMSTASAGLLAASTTVTE
                     DLLPKLRGGKQSSLGTSRLFTLLTGLVVLGIALMVNDVINALTLAYNLLVGGMLIPLI
                     GAIFWKRATTAGAIASMSLGFATALLFMFLDGLEANTPIYYSLAVGLVSFVGVSLVSR
                     KSVAAVKLA"
     gene            complement(9174. .10067)
                     /locus_tag="PP4525"
     CDS             complement(9174. .10067)
                     /locus_tag="PP4525"
                     /note="similar to GP:9947361; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, LysR family"
                     /protein_id="AAN70099.1"
                     /db_xref="GI:24986259"
                     /translation="MEFSQLRIFQAVAEEGSVTRAAERLHRVPSNLSTRLRQLEEQLG
                     VELFLRERQRLQLSPAGKVLLDYANRMSALRDEALGAVRGGQPAGDFVLGTMYSTAAT
                     HLPALLARYHQSYPQVNLQVRAAPSGELLEGLLNHTLDAALVDGPPSLAGLDGVPLCD
                     EQLVLITSPEHPAVLTAKDVAGKAVFTFRQGCSYRMRLEAWYAHAHTPMGRVMEIESY
                     QSMLACVIAGAGVAMMAQSMLDSLPGRDRVRVHSLQAPFDQAVTWLMWRQGMRGANLQ
                     AWIDLQQSETINQSSECAATA"
     gene            10172. .11344

  Query Match             32.9%;  Score 296;  DB 1;  Length 300933;
  Best Local Similarity   59.6%;  Pred. No. 3.8e-21;
  Matches  553;  Conservative    0;  Mismatches  345;  Indels   30;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || |||||||| |||||||||| |  ||||||||| | |||||||   ||
Db     206468 GGGATGCCCAGGCGCTGGCGGCGTTCCCACAGGCACTTGCGGCTTACGCCGAGCTTGCGT 206409

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  | |||||||||| || |||||||||  |  |||| ||| |

Db     206408 GCCAGTTCGGTCTCGGTCATGTGGTCCTGGTGTTCAAGCACGAAGTGCTGGAAGTAGTCT 206349

Qy        123 TCGAGCGA---GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGA 179
              || |||||    |||    | |||| |   |  |           |||  ||  ||   | |
Db     206348 TCCAGCGACAGGTCTTCGGTCGGCTCGTGGCTGGCATTGCCCGCGTTGGCCACCACCAGG 206289

Qy        180 GGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCC------- 232
              |      |  ||| ||    | |||| |  |  ||||||  || ||||
Db     206288 GCATTGTCGAGGATGTCATCCTCATCCAGGTCGCTCAGCTCGATGTCGATGCCCAGCAGC 206229

Qy        233 --------------------CCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
                                   | ||  ||| | ||| ||||| |  |||||||| |||||
Db     206228 TCGGCGGAAATTTCTGCGCTCTCGCTGAGGATCACCGCGCGCTCCACGGCGTTTTCCAGT 206169

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || || ||||| || |||||              |||||       ||||| ||||| |
Db     206168 TCACGCACGTTACCGGGCCAGCTGTAGTGACGAATGGCTTGCTCGGCCTCGGCCGAGAAG 206109

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
                ||    |  | || ||        || | ||   ||  |   | ||||||
Db     206108 TGCAGGTCGTCGCGGCCGATGCGAGCGCTCTGGCGGGCGAGGAAGACGTTGGCGATCTCG 206049

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
                |||| | |   ||||| ||||| | |||||||    |   | |  ||  |||| |||
Db     206048 TTGACGTCGCTGCCACGCTCGCGCAGGGCCGGCAGCTTCAGGGCGATGACGTGCAGTCGG 205989

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
               |||||| |||    |||||    ||  |     |||         || || ||||| |||
Db     205988 TAATACAAGTCTTCACGGAACTGGCCGGCCTTGGCCAGGTTTTTCAGGTCGCGGTGGGTC 205929

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || |||||||| ||||| ||  ||   | |   ||  ||| || ||||| || |  |||
Db     205928 GCGGCGATCAGGCGCACATCGACCTTTTGCGACTGCACCGAACCTACCCGGCGTATCTCG 205869

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||    |||||  | ||||||||||||  ||||  | |  | |||||| ||||| |  ||
Db     205868 CCTTCCTGCAGCACACGCAGCAGGCGAGCCTGGGCTTCGAGCGGCAGTTCGCCAATTTCA 205809

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              ||||| | |||||| ||||| || || |||||    ||  || || ||| | |||   |||
Db     205808 TCGAGGAACAGCGTACCGCCATCGGCGGCCTCCACCAAGCCGGCGCGCCCGGCGCTGGCG 205749

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
              || | |  ||||||    ||||  |||||||| |||   |||||| |||  ||||||||
Db     205748 CCGGTGAACGCGCCTTTTTCGTGGCCGAACAGTTCGGATTCGATCAGCGTTTCGGGGATG 205689

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              |||||||| | |||    | | |||   ||    |||  |  |||  |||||  |||
Db     205688 GCCGCGCAGTTCACCGAGATCATCGGTGCCTTGGCCCGGCGAGACAGGTTGTGCAGGGCG 205629

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || |||| |  ||| || || ||||||  |||||||    | |||||    | |  ||| |
Db     205628 CGGGCCACCAGCTCTTTACCGGTGCCCGACTCGCCCTGGATCAGCACATTGGAGTCGGTA 205569

Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||| |||||||| | ||    | |||
Db     205568 GGCGCGACCTTGCGGATCTTGCTGTACA 205541


RESULT 14
CP000076_60/c
WPCOMMENT
  Sequence split into 71 fragments  LOCUS CP000076 Accession CP000076
     Fragment Name      Begin      End
     CP000076_00            1    110000
     CP000076_01       100001    210000
     CP000076_02       200001    310000
     CP000076_03       300001    410000
     CP000076_04       400001    510000
     CP000076_05       500001    610000
     CP000076_06       600001    710000
     CP000076_07       700001    810000
     CP000076_08       800001    910000
```

CP000076_09	900001	1010000
CP000076_10	1000001	1110000
CP000076_11	1100001	1210000
CP000076_12	1200001	1310000
CP000076_13	1300001	1410000
CP000076_14	1400001	1510000
CP000076_15	1500001	1610000
CP000076_16	1600001	1710000
CP000076_17	1700001	1810000
CP000076_18	1800001	1910000
CP000076_19	1900001	2010000
CP000076_20	2000001	2110000
CP000076_21	2100001	2210000
CP000076_22	2200001	2310000
CP000076_23	2300001	2410000
CP000076_24	2400001	2510000
CP000076_25	2500001	2610000
CP000076_26	2600001	2710000
CP000076_27	2700001	2810000
CP000076_28	2800001	2910000
CP000076_29	2900001	3010000
CP000076_30	3000001	3110000
CP000076_31	3100001	3210000
CP000076_32	3200001	3310000
CP000076_33	3300001	3410000
CP000076_34	3400001	3510000
CP000076_35	3500001	3610000
CP000076_36	3600001	3710000
CP000076_37	3700001	3810000
CP000076_38	3800001	3910000
CP000076_39	3900001	4010000
CP000076_40	4000001	4110000
CP000076_41	4100001	4210000
CP000076_42	4200001	4310000
CP000076_43	4300001	4410000
CP000076_44	4400001	4510000
CP000076_45	4500001	4610000
CP000076_46	4600001	4710000
CP000076_47	4700001	4810000
CP000076_48	4800001	4910000
CP000076_49	4900001	5010000
CP000076_50	5000001	5110000
CP000076_51	5100001	5210000
CP000076_52	5200001	5310000
CP000076_53	5300001	5410000
CP000076_54	5400001	5510000
CP000076_55	5500001	5610000
CP000076_56	5600001	5710000
CP000076_57	5700001	5810000
CP000076_58	5800001	5910000
CP000076_59	5900001	6010000
CP000076_60	6000001	6110000
CP000076_61	6100001	6210000
CP000076_62	6200001	6310000
CP000076_63	6300001	6410000
CP000076_64	6400001	6510000
CP000076_65	6500001	6610000
CP000076_66	6600001	6710000
CP000076_67	6700001	6810000
CP000076_68	6800001	6910000
CP000076_69	6900001	7010000
CP000076_70	7000001	7074893

Continuation (61 of 71) of CP000076 from base 6000001 (CP000076 Pseudomonas fluorescens

Query Match 31.4%; Score 283; DB 1; Length 110000;
Best Local Similarity 58.0%; Pred. No. 9.7e-20;
Matches 533; Conservative 0; Mismatches 365; Indels 21; Gaps 1;

```
Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || |||||||||| |  |||||||||| ||| ||    ||
Db      40728 GGGATGCCCAGGCGCTGACGACGTTCCCACAGGCACTTGCGGCTGACGCCCAGTTTGCGT 40669

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
```

```
                 || | ||||||||| ||||  | ||||||||| || || |||||| |  |||| ||| ||
Db      40668 GCCAGCTCGGTCTCGGTCATGTGGTCCTGGTGTTCGAGGACGAAGTGCTGGAAGTAGTCC 40609

Qy        123 TCGAGCGAGTCCGAAGGCGGCGCCCCGTCGC--------------------GCAGCGAT 161
              || || ||          | ||   ||| ||                    ||||
Db      40608 TCCAGGGACAGGTCTTCGGTCGGTTCGTGGCTGGTGTTACCCGCTTGTGGGGGCAGGCCG 40549

Qy        162 GCGGAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCG 221
              || ||  | ||        |    | | |   | |||||  || ||  || |   | |
Db      40548 ACGAAGTCGTCGTCTTCCAGGTCGCTCAGTTCGATGTCGATGCCCAGCAGGTCGGCCGAG 40489

Qy        222 ATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACG 281
              ||  |  ||  ||  || |  || || || |  || ||||||||||| || || |||
Db      40488 ATTTCCGGGCTTTCGCACAGAATGACCGCACGCTCGACAGCGTTCTCCAGTTCGCGCACG 40429

Qy        282 TTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGC 341
              || |||||||| |          |||||||     || || || |  |  ||   |
Db      40428 TTCCCCGGCCAGGAGTAATGGCGGATGGCCTGTTCTGCATCGGCGGCGAACTTAAGGTCC 40369

Qy        342 GCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCC 401
              |   | ||  |    |   || | ||  ||| |  |  | ||| ||       ||||
Db      40368 GTGCGGCCGACGCGCGCGCTCTGGCGGGCCAGAAAGGCATTGGCAATTTCATTGACGTCT 40309

Qy        402 CCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGG 461
              |  |||||  |||||  | ||||||    |  | | |||  | || ||| | ||||||
Db      40308 GCACCGCGCTCACGCAGGGCCGGCAGTTTCAGGGCAATCACGTGGAGGCGGTAGTACAGG 40249

Qy        462 TCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATC 521
              ||   |||||   ||      |  ||         || || ||||| || || || |||
Db      40248 TCTTCACGGAACTGGCCTATTTTCGACAGGCTTTTCAGGTCGCGGTGAGTGGCTGCAATC 40189

Qy        522 AGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGC 581
              || ||||| ||  |   | | |   | ||| || || || || |  |||||   ||||
Db      40188 AGGCGCACATCGACTTTCTGTGACTGTACCGAGCCGACACGGCGGATCTCGCCTTCCTGC 40129

Qy        582 AAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGC 641
              |  |  |||||||  ||| ||||| ||| || ||| |||||||| | |||||||| | |
Db      40128 AGTACTCGCAGCAATCGCGCCTGCGCCTCCAGCGGTAGCTCGCCAATCTCGTCGAGGAAC 40069

Qy        642 AGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCC 701
              || || ||||| ||||| || ||   ||| || || ||  | |||   || || | | |
Db      40068 AGGGTGCCGCCGTCCGCAGCTTCTACCAGGCCGGCACGGCCGGCGCTGGCCCCGGTGAAC 40009

Qy        702 GCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCAC 761
              |||||   |||||  |||||||| |||  ||||||| | |  || || || || |||||
Db      40008 GCGCCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTTTCCGGAATGGCTGCGCAG 39949

Qy        762 TGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGC 821
              | |||    |  | |||   ||     |||  |  |||   |||||  || || || | |
Db      39948 TTCACCGAAATCATCGGTGCCTTGGCACGCTTGGACAGGTTGTGCAGGGCCCGGGCTACC 39889

Qy        822 GTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACC 881
              |||||||| || ||    || |||   | ||| |   | |  ||  | ||||||||
Db      39888 AGTTCCTTGCCGGTACCGGATTCACCCTGAATCAGTACATTGGAGTCAGTGGGCGCCACT 39829

Qy        882 TTGCGCACCTGCGCGAACA 900
              ||||| | ||    | |||
Db      39828 TTGCGGATCTTGCTGTACA 39810
```

RESULT 15
CP000075_09
WPCOMMENT
Sequence split into 61 fragments LOCUS CP000075 Accession CP000075

Fragment Name	Begin	End
CP000075_00	1	110000
CP000075_01	100001	210000
CP000075_02	200001	310000
CP000075_03	300001	410000
CP000075_04	400001	510000
CP000075_05	500001	610000
CP000075_06	600001	710000
CP000075_07	700001	810000

CP000075_08	800001	910000
CP000075_09	900001	1010000
CP000075_10	1000001	1110000
CP000075_11	1100001	1210000
CP000075_12	1200001	1310000
CP000075_13	1300001	1410000
CP000075_14	1400001	1510000
CP000075_15	1500001	1610000
CP000075_16	1600001	1710000
CP000075_17	1700001	1810000
CP000075_18	1800001	1910000
CP000075_19	1900001	2010000
CP000075_20	2000001	2110000
CP000075_21	2100001	2210000
CP000075_22	2200001	2310000
CP000075_23	2300001	2410000
CP000075_24	2400001	2510000
CP000075_25	2500001	2610000
CP000075_26	2600001	2710000
CP000075_27	2700001	2810000
CP000075_28	2800001	2910000
CP000075_29	2900001	3010000
CP000075_30	3000001	3110000
CP000075_31	3100001	3210000
CP000075_32	3200001	3310000
CP000075_33	3300001	3410000
CP000075_34	3400001	3510000
CP000075_35	3500001	3610000
CP000075_36	3600001	3710000
CP000075_37	3700001	3810000
CP000075_38	3800001	3910000
CP000075_39	3900001	4010000
CP000075_40	4000001	4110000
CP000075_41	4100001	4210000
CP000075_42	4200001	4310000
CP000075_43	4300001	4410000
CP000075_44	4400001	4510000
CP000075_45	4500001	4610000
CP000075_46	4600001	4710000
CP000075_47	4700001	4810000
CP000075_48	4800001	4910000
CP000075_49	4900001	5010000
CP000075_50	5000001	5110000
CP000075_51	5100001	5210000
CP000075_52	5200001	5310000
CP000075_53	5300001	5410000
CP000075_54	5400001	5510000
CP000075_55	5500001	5610000
CP000075_56	5600001	5710000
CP000075_57	5700001	5810000
CP000075_58	5800001	5910000
CP000075_59	5900001	6010000
CP000075_60	6000001	6093698

Continuation (10 of 61) of CP000075 from base 900001 (CP000075 Pseudomonas syringae pv.

Query Match 31.4%; Score 282.6; DB 1; Length 110000;
Best Local Similarity 58.2%; Pred. No. 1.1e-19;
Matches 535; Conservative 0; Mismatches 354; Indels 30; Gaps 1;

```
Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || ||||||||||      ||||| |||| ||| |||   ||
Db      41351 GGGATGCCCAGACGCTGACGGCGTTCCCACAGACATTTGCGACTGACGCCCAGCTTGCGC 41410

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || | |||||| || ||||  |  ||||| |||||||| |||||| |   | |||||| |
Db      41411 GCCAGCTCGGTTTCAGTCATGTGATCCTGATGCTCCAGAACGAAGTGCTGAAAATAGTCT 41470

Qy        123 TCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAGGC 182
              ||||| ||          |  | | ||| || |  |||  |||| ||| | ||    |||
Db      41471 TCGAGTGACAGGTCTTCGGCGGGCTCGTTGCCCGACGAGCCGGAAGAGCCCGGTAACGGC 41530

Qy        183 GGC-----------------------------CGCGGCGGGTCGTCGAGCCCGAGGTGG 212
              | |                             ||  |  |||||  || ||  ||
```

```
Db      41531 GCCATGCCCATGTACTCGTCGTCATCCAGACCGTCGAGCTCGATGTCGATGCCCAGCAGG 41590

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
              |       |||  |   |  | ||   || | ||| | ||| |  || |||||  ||||||
Db      41591 TCAGCGGTGATTTCCGGGTTCTCGCACAGAATCACCGAGCGTTCCACAGCGTTTTCCAGC 41650

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || || ||||| ||||||||              ||| ||        || ||    |  |
Db      41651 TCGCGTACGTTACCCGGCCAGCTGTAATGGCGGATCGCTTGTTCGGCGTCGTGTGCAAAT 41710

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
              |||     |  |||| | |   |   || | ||  ||| | ||  | ||||||  |
Db      41711 TTGAGGTCGTCGCGCCCGACCTTGCCGCTCTGGCGAACCAGGAACGCATGGGCGATTTCC 41770

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
              |    |||   |   || || ||||||| ||| |||   |   | |  ||  |||| ||
Db      41771 AGGATGTCGTTGCCACGTTCGCGCAGCGCCGGGAGCTTCAGGGCGATGACGTGCAGACGA 41830

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
              | |||||||    || ||    || |      ||| ||      || ||||||| |||
Db      41831 TAGAACAGGTCTTCACGAAACTCGCCATTTTTGGCCAGCGTTTTCAGGTCCCGGTGGGTC 41890

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || || ||||| || |||||  ||    | |   ||  ||| || ||||| || |  |||
Db      41891 GCGGCAATCAGGCGTACGTCGACCTTTTGCGATTGCACCGAACCGACCCGGCGAATCTCG 41950

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||   |||||| |  ||||||||||  ||||  | | || ||||||||||||||  ||
Db      41951 CCTTCCTGCAATACGCGCAGCAGGCGAGCCTGGGCTTCCAGCGGCAGCTCGCCGATTTCA 42010

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              ||||| | |||||| || || || || ||||||  ||| || || ||  | |||   ||
Db      42011 TCGAGAAACAGCGTGCCACCGTCAGCAGCCTCGACCAGACCTGCGCGGCCTGCGCTGGCC 42070

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
              || | |  ||||||    || |  |||||||| |||  ||||||| |||  ||||| ||
Db      42071 CCGGTGAACGCGCCTTTTTCATGACCGAACAGTTCGGACTCGATCAGCGATTCGGGAATG 42130

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              ||||| || |  |||    |  | ||||  ||       ||  |     ||  |||||  |||
Db      42131 GCCGCACAGTTGACGGAAATCATCGGCGCCTTGGCGCGGCGTGAGAGGTTGTGCAGGGCG 42190

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || || | |   ||||| || ||||||   || ||     | ||| |   | |  ||  ||
Db      42191 CGGGCGACCAGTTCCTTACCGGTGCCCGATTCCCCTTGAATCAGGACGTTGGAATCAGTC 42250

Qy        873 GGCGCCACCTTGCGCACCT 891
              || || || ||||| | ||
Db      42251 GGGGCAACTTTGCGGATCT 42269
```

Search completed: April 28, 2006, 02:39:27
Job time : 4991.97 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 433.991 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_1_900
Perfect score: 900
Sequence: 1 tgggtatcccgagccgctgg.....cttgcgcacctgcgcgaaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	900	100.0	76994	12	ADP64454	Adp64454 Sorangium
	2	336	37.3	1077	11	ABD12115	Abd12115 Pseudomon
c	3	336	37.3	1173	11	ABD11844	Abd11844 Pseudomon
c	4	327.6	36.4	1416	2	AAV20099	Aav20099 Pseudomon
c	5	327.6	36.4	1416	3	AAA13891	Aaa13891 Pseudomon
c	6	327.6	36.4	1416	6	AAD22868	Aad22868 Pseudomon
c	7	327.6	36.4	4377	2	AAV23493	Aav23493 Pseudomon
c	8	327.6	36.4	4377	3	AAA13904	Aaa13904 Pseudomon
c	9	327.6	36.4	4377	4	AAF30869	Aaf30869 Pseudomon
c	10	327.6	36.4	4377	6	AAD22881	Aad22881 P. alcali
	11	274.8	30.5	1386	11	ABD16927	Abd16927 Pseudomon
c	12	274.8	30.5	1461	11	ABD16704	Abd16704 Pseudomon
c	13	267.8	29.8	972	14	ACL72552	Acl72552 M. xanthu
c	14	267.8	29.8	1053	14	ACL69201	Acl69201 M. xanthu
	15	267.8	29.8	11812	14	ACL64578	Acl64578 M. xanthu
	16	267.8	29.8	12849	14	ACL64500	Acl64500 M. xanthu
c	17	266	29.6	2379	11	ABD10601	Abd10601 Pseudomon
c	18	264.6	29.4	1617	14	ACL68921	Acl68921 M. xanthu
	19	264.6	29.4	19053	3	AAA59146	Aaa59146 Nucleotid
	20	264.6	29.4	19053	13	ADU66531	Adu66531 Myxococcu
	21	264.6	29.4	19053	14	ADZ44588	Adz44588 Myxococcu
c	22	264.6	29.4	34552	14	ACL64799	Acl64799 M. xanthu
c	23	264.4	29.4	1131	8	ACA42322	Aca42322 Prokaryot
c	24	260.8	29.0	837	11	ABD10519	Abd10519 Pseudomon
c	25	260	28.9	1362	14	ACL68691	Acl68691 M. xanthu
c	26	260	28.9	29559	14	ACL64791	Acl64791 M. xanthu
c	27	256.2	28.5	1344	14	ACL71095	Acl71095 M. xanthu
	28	256.2	28.5	6155	14	ACL64274	Acl64274 M. xanthu
c	29	254.6	28.3	3699	14	ACL67268	Acl67268 M. xanthu
	30	254.6	28.3	19112	14	ACL64718	Acl64718 M. xanthu
c	31	250.4	27.8	1107	14	ACL71787	Acl71787 M. xanthu
	32	250.4	27.8	6698	14	ACL64389	Acl64389 M. xanthu
c	33	248.8	27.6	1356	14	ACL72446	Acl72446 M. xanthu
	34	248.8	27.6	9091	14	ACL64486	Acl64486 M. xanthu
c	35	247.8	27.5	1350	14	ACL68082	Acl68082 M. xanthu
c	36	247.8	27.5	10835	14	ACL64568	Acl64568 M. xanthu
c	37	245.6	27.3	1422	14	ACL66656	Acl66656 M. xanthu
c	38	245.6	27.3	15268	14	ACL64679	Acl64679 M. xanthu
c	39	245.2	27.2	1341	14	ACL72363	Acl72363 M. xanthu
c	40	245.2	27.2	8765	14	ACL64476	Acl64476 M. xanthu
c	41	244.8	27.2	1434	14	ACL68747	Acl68747 M. xanthu
c	42	244.8	27.2	31826	14	ACL64793	Acl64793 M. xanthu
c	43	244.2	27.1	1491	14	ACL66872	Acl66872 M. xanthu
c	44	244.2	27.1	16387	14	ACL64693	Acl64693 M. xanthu
c	45	243.4	27.0	1252	14	ACL70286	Acl70286 M. xanthu

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

```
PS   Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC   The present invention describes an isolated recombinant polynucleotide
CC   (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC   synthase (PKS) protein or a fragment comprising one or more domains of
CC   the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC   host cell (III) comprising (II), or (I) integrated into the cell
CC   chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC   disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC   the native disorazole PKS by the inactivation of one or more domain of
CC   the disorazole PKS and/or addition of one or more domain of a non-
CC   disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC   expression system (VI) capable of producing a disorazole synthase domain
CC   in a host cell, where the system comprises an encoding sequence for a
CC   disorazole polyketide synthase domain, and the encoding sequence is
CC   operably linked to control sequences effective in the cell to produce RNA
CC   that is translated into the domain; (7) host cell modified to contain
CC   (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC   in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC   eliminate production of disorazole; (9) an isolated polypeptide encoded
CC   by (I); (10) hybrid polyketide synthase comprising one or more
CC   polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC   disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC   of at least 200 base pairs with a sequence identical or substantially
CC   identical to a protein encoding region of the 77294 base pair sequence of
CC   SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC   inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC   (I) is useful for producing disorazole polyketides. (III) is useful for
CC   producing a polyketide, which involves growing (III) under conditions,
CC   where a polyketide synthesised by a PKS comprising a protein encoded by
CC   the polynucleotide molecule is produced in the cell. (I) is useful for
CC   producing libraries of PKSs, where the polyketide producing colonies are
CC   identified and isolated, and the produced polyketides are used
CC   collectively in a panel to represent a library or assessed individually
CC   for activity. The disorazole polyketide produced by (I) is useful in
CC   molecular biology, chemistry, recombinant DNA technology, human and
CC   veterinary medicine and agricultural applications. The disorazole
CC   polyketide produced by (I) is formulated as a pharmaceutical composition,
CC   useful for its activity against microorganisms, in a mammal. The present
CC   sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC   disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 900;  DB 12;  Length 76994;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-111;
  Matches  900;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGGGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGGGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGAC 60

Qy         61 GGGCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGC 120

Qy        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180

Qy        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240

Qy        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300

Qy        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360

Qy        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCG 420

Qy        421 GCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCT 480

Qy        481 CGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCC 540

Qy        541 GGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCC 600

Qy        601 CCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCG 660

Qy        661 CCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGA 720

Qy        721 ACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCT 780

Qy        781 GCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCC 840

Qy        841 CCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900


RESULT 2
ABD12115
ID   ABD12115 standard; DNA; 1077 BP.
XX
AC   ABD12115;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #10719.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO78544.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 10719; 455pp; English.
```

```
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 1077 BP; 142 A; 348 C; 384 G; 203 T; 0 U; 0 Other;

  Query Match             37.3%;  Score 336;  DB 11;  Length 1077;
  Best Local Similarity   62.3%;  Pred. No. 3.1e-36;
  Matches  578;  Conservative    0;  Mismatches  320;  Indels   30;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || |||||||| |||||||||| |   ||||||||||||||| |||    |||
Db         11 GGAATGCCCAGGCGCTGGCGGCGTTCCCACAGGCATTTGCGGCTGATGCCCAGCTTGCGG 70

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || | |||||||||| ||||  | ||||||||||||| || ||||||  |  |||| ||| |
Db         71 GCCAACTCGGTCTCGGTCATGTGGTCCTGGTGCTCGAGGACGAAGTGCTGGAAGTAGTCT 130

Qy        123 TCGAGCGA--------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATG 162
              || |||||                   |   |  ||| ||| | |    ||   ||| |
Db        131 TCCAGCGAAAGGTCCTCGGTGGGCTCGTGGTTGCTGGCCGCGGCGCCCGCCGGGCCGAGG 190

Qy        163 CGGAGGAGACGGGCGGAGGCGGCCGCG----------GCGGGTCGTCGAGCCCGAGGTGG 212
              | ||||     |||| || || || ||          ||  |  |||||  || ||  ||
Db        191 CAGAGGTCGGCGGCGAAGTCGTCCTCGAGGTCGTCCAGCTCGATGTCGATACCCAGCAGG 250

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
                |   | |||  |   || | ||  ||| | ||| || || |  | ||||||  |||||
Db        251 TCGGCGTGGATTTCCTGGCTCTCGCTGAGGATCACCGCCCGCTCGATGGCGTTTTCCAGT 310

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || ||  |||||||||||||| ||         ||| |||     || ||    |  |
Db        311 TCGCGCACGTTGCCCGGCCATGGGTAATGGCGGATCGCCTGTTCGGCGTCGTGGGCGAAG 370

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
              |||||||||  ||||||||          | | ||   ||| |  |    ||||||   |
Db        371 CTGAGCGGCGCGCGCCCCATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCC 430

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
              | | ||||   |   |||||  |||||  |  ||||||   | |    | |||  |||| |||
Db        431 ATCACGTCGTTGCCACGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGG 490

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
              ||||  |||||    |||||   |||         |||  |      || || ||||| ||
Db        491 TAATAGAGGTCTTCGCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTG 550

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || |||||||| |||||||| ||  | | |   ||  ||| || ||||| || |   |||
Db        551 GCGGCGATCAGGCGCACGTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCG 610

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||   |||||| |  |||||||||||  ||||  ||| || |||||| ||||||| ||||
Db        611 CCTTCCTGCAATACGCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCG 670

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              ||||| | ||| || ||||| || ||||| |||  |||||| || ||| | |||   ||
Db        671 TCGAGGAACAGGGTGCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCA 730

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
```

```
              || | |   |||||    || |  ||||||||||||   |||||| | |  || || |||
Db        731 CCGGTGAAGGCGCCTTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTCCGGAATC 790

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              ||||| || | |||     |  | ||||  ||      |||      |||  ||||| |||
Db        791 GCCGCACAGTTCACCGAGATCAGCGGCGCCTTGGCGCGCTTCGACAGGTTGTGCAGGGCG 850

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || || | |   |||||||| |||||   |||||||   | ||| | ||| |  ||| ||
Db        851 CGGGCGACCAGTTCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACCGTGGAGTCGGTC 910

Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||| |||||||| | ||    | |||
Db        911 GGCGCGACCTTGCGAATCTTGCTGTACA 938


RESULT 3
ABD11844/c
ID   ABD11844 standard; DNA; 1173 BP.
XX
AC   ABD11844;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #10448.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO78273.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 10448; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 1173 BP; 222 A; 416 C; 378 G; 157 T; 0 U; 0 Other;

  Query Match             37.3%;  Score 336;  DB 11;  Length 1173;
```

```
  Best Local Similarity   62.3%;  Pred. No. 3.1e-36;
  Matches  578;  Conservative    0;  Mismatches  320;  Indels   30;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || |||||||| |||||||||  |   |||||||||||||| |||   |||
Db       1139 GGAATGCCCAGGCGCTGGCGGCGTTCCCACAGGCATTTGCGGCTGATGCCCAGCTTGCGG 1080

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || | |||||||||  ||||  | ||||||||||||  || ||||||  |  ||||  ||| |
Db       1079 GCCAACTCGGTCTCGGTCATGTGGTCCTGGTGCTCGAGGACGAAGTGCTGGAAGTAGTCT 1020

Qy        123 TCGAGCGA-------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATG 162
              || |||||                   |   |  ||| ||| | |    ||   ||| |
Db       1019 TCCAGCGAAAGGTCCTCGGTGGGCTCGTGGTTGCTGGCCGCGGCGCCCGCCGGGCCGAGG 960

Qy        163 CGGAGGAGACGGGCGGAGGCGGCCGCG----------GCGGGTCGTCGAGCCCGAGGTGG 212
              | ||||     |||| || || || ||          ||  |  |||||  || ||  ||
Db        959 CAGAGGTCGGCGGCGAAGTCGTCCTCGAGGTCGTCCAGCTCGATGTCGATACCCAGCAGG 900

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
              |    | |||  |   || | ||  ||| | ||| || || |  | ||||| |||||
Db        899 TCGGCGTGGATTTCCTGGCTCTCGCTGAGGATCACCGCCCGCTCGATGGCGTTTTCCAGT 840

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
              || || |||||||||||||| ||          ||| |||      || ||    |  |
Db        839 TCGCGCACGTTGCCCGGCCATGGGTAATGGCGGATCGCCTGTTCGGCGTCGTGGGCGAAG 780

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392
                |||||||| |||||||         | | ||   ||| |  |    |||||  |
Db        779 CTGAGCGGCGCGCGCCCCATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCC 720

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452
              | | ||||   |   ||||| ||||| |  ||||||   | |   | |||  |||| |||
Db        719 ATCACGTCGTTGCCACGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGG 660

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512
              ||||  |||||    |||||   |||         |||  |      || || ||||| ||
Db        659 TAATAGAGGTCTTCGCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTG 600

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572
              || |||||||| |||||||  ||  | | |   ||  ||| || ||||| || |  |||
Db        599 GCGGCGATCAGGCGCACGTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCG 540

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632
              ||    |||||| |  |||||||||||  ||||  ||| || |||||| |||||| ||||
Db        539 CCTTCCTGCAATACGCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCG 480

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692
              |||||  | ||| || ||||| || |||||  |||  |||||| || ||| | |||   ||
Db        479 TCGAGGAACAGGGTGCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCA 420

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752
              || | |   |||||   || |  |||||||||||  ||||||  | |   || || |||
Db        419 CCGGTGAAGGCGCCTTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTCCGGAATC 360

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812
              ||||| || | |||     |  | ||||  ||      |||    |||  |||||  |||
Db        359 GCCGCACAGTTCACCGAGATCAGCGGCGCCTTGGCGCGCTTCGACAGGTTGTGCAGGGCG 300

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872
              || ||  | |   |||||||| |||||   |||||||    | |||  | |||  |||  ||
Db        299 CGGGCGACCAGTTCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACCGTGGAGTCGGTC 240

Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900
              |||||  |||||||| | ||    | |||
Db        239 GGCGCGACCTTGCGAATCTTGCTGTACA 212


RESULT 4
AAV20099/c
ID   AAV20099 standard; DNA; 1416 BP.
XX
AC   AAV20099;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Pseudomonas LipR kinase coding sequence.
XX
KW   Kinase; LipQ; LipR; lipase expression regulator; DNA binding regulator;
KW   sigma 54 promoter; secretion factor; lux-box binding element;
KW   orfV-box binding element; regulation cascade; ss.
XX
OS   Pseudomonas alcaligenes.
XX
PN   WO9806836-A2.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US014450.
XX
PR   16-AUG-1996;   96US-00699092.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 1998-159528/14.
DR   P-PSDB; AAW53825.
XX
PT   Nucleic acids encoding proteins involved in the lipase regulation cascade
PT   from P. alcaligenes - useful for controlling production and secretion of
PT   heterologous proteins in P. alcaligenes.
XX
PS   Claim 7; Page 23; 106pp; English.
XX
CC   This sequence encodes the LipR kinase of Pseudomonas alcaligenes. The DNA
CC   represents a nucleic acid of the invention. The nucleic acids encode: (a)
CC   a kinase from a Pseudomonad that regulates the expression of a lipase;
CC   (b) a DNA binding regulator from a Pseudomonad that regulates the
CC   expression of a lipase; (c) a Pseudomonas alcaligenes upstream activating
CC   sequence; (d) a P. alcaligenes sigma 54 promoter that regulates the
CC   expression of a lipase; (e) a P. alcaligenes secretion factor selected
CC   from XcpP, Q, R, S, T, U, V, W, X, Y and Z, and OrfV, X, Y; (f) a P.
CC   alcaligenes lux-box binding element; and (g) a orfV-box binding element.
CC   The nucleic acids represent parts of a regulation cascade, comprising at
CC   its heart a kinase and a DNA binding regulator. These sequences can be
CC   used for the production of heterologous proteins in a host cell. The
CC   cascade also comprises secretion factors which can enhance the secretion
CC   of produced proteins
XX
SQ   Sequence 1416 BP; 258 A; 510 C; 453 G; 195 T; 0 U; 0 Other;

  Query Match            36.4%;  Score 327.6;  DB 2;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 3.8e-35;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || |||||||  |  ||||||||||||||||||||    ||
Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| ||||||  |  ||||  ||| ||
Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                 ||     | ||  | |    |  || | |    ||
Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |     |       | ||  |  |||||  || ||  |  |        ||
Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| ||||| |  | ||| ||||||||||| || ||||||
Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083
```

```
Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900
Db        482 CGGATCTTGCCGTAAA 467
```

```
RESULT 5
AAA13891/c
ID   AAA13891 standard; DNA; 1416 BP.
XX
AC   AAA13891;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Pseudomonas alcaligenes LipR nucleotide sequence SEQ ID NO:3.
XX
KW   Pseudomonas alcaligenes; expression; lipase regulation cascade; kinase;
KW   DNA binding regulator; polymerase; promoter; secretion factor; XcpP;
KW   XcpQ; XcpR; XcpS; XcpT; XcpU; XcpV; XcpW; XcpX; XcpY; XcpZ; OrfV; OrfX;
KW   OrfY; OrfZ; LipQ; LipR; upstream activating sequence; detergent;
KW   cleaning formulation; ss.
XX
OS   Pseudomonas alcaligenes.
XX
PN   US6048710-A.
XX
PD   11-APR-2000.
XX
PF   15-AUG-1997;   97US-00911853.
XX
PR   16-AUG-1996;   96US-00699092.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 2000-316896/27.
DR   P-PSDB; AAY82589.
XX
PT   Expression vector for producing heterologous proteins in host cells
PT   comprises a nucleic acid encoding a kinase and a DNA binding regulator
PT   which hybridizes under stringent conditions to nucleic acid.
XX
PS   Claim 1; Fig 2; 133pp; English.
XX
CC   The present invention describes an expression vector comprising a nucleic
CC   acid encoding a kinase and a DNA binding regulator which hybridises under
CC   stringent conditions to a nucleic acid isolated from Pseudomonas
CC   alcaligenes. Also described are: (1) an isolated plasmid comprising the
CC   above expression vector; (2) a method of transforming a host cell
CC   comprising adding the above plasmid to host cells under appropriate
CC   conditions; (3) a transformed host cell comprising the above expression
CC   vector; and (4) a method for producing a protein comprising the steps of
CC   obtaining a host cell comprising the above expression vector and further
CC   comprising nucleic acid encoding the protein, and culturing the host cell
CC   under conditions for the expression of protein. The expression vector of
CC   the present invention can be used for producing heterologous proteins in
CC   host cells, particularly, lipase in Pseudomonas. Lipases produced can be
CC   used in detergents and cleaning formulations in industrial processes. The
CC   invention provides a higher production level and efficiently express a
CC   heterologous protein. The present sequence encodes LipR isolated from
CC   Pseudomonas alcaligenes, from the present invention
XX
SQ   Sequence 1416 BP; 258 A; 510 C; 453 G; 195 T; 0 U; 0 Other;
```

```
  Query Match           36.4%;  Score 327.6;  DB 3;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 3.8e-35;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
```

```
Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644

Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704

Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764

Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824

Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884

Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900

Db        482 CGGATCTTGCCGTAAA 467


RESULT 6
AAD22868/c
ID   AAD22868 standard; DNA; 1416 BP.
XX
AC   AAD22868;
XX
DT   26-FEB-2002  (first entry)
XX
DE   Pseudomonas alcaligenes LipR DNA.
XX
KW   DNA binding regulator; lipase regulation cascade; kinase; polymerase;
KW   promoter; upstream activating factor; secretion factor; detergent;
KW   cleaning formulation; industrial process; LipR; ds.
XX
OS   Pseudomonas alcaligenes.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1416
FT                   /*tag=  a
FT                   /product= "LipR protein"
XX
PN   US6313283-B1.
XX
PD   06-NOV-2001.
XX
PF   06-JAN-2000; 2000US-00479453.
XX
PR   16-AUG-1996;   96US-00699092.
PR   15-AUG-1997;   97US-00911853.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 2002-024912/03.
DR   P-PSDB; AAE13655.
XX
PT   Polynucleotide encoding DNA binding regulator, useful for regulating
PT   expression of lipase especially in pseudomonad, and lipase useful in
PT   detergents and other cleaning formulations and in various industrial
PT   processes.
XX
PS   Claim 1; Fig 2; 132pp; English.
XX
CC   The present invention relates to a polynucleotide encoding a DNA binding
CC   regulator that can regulate the expression of a lipase. The invention
CC   also relates to an expression system comprising components of Pseudomonas
CC   alcaligenes lipase regulation cascade which includes kinases, DNA binding
CC   regulators, polymerases, promoters, upstream activating factors and
CC   secretion factors. DNA binding regulators of the invention are useful for
CC   regulating the expression of a lipase, where the lipase is useful in
CC   detergents and other cleaning formulations as well as a number of
CC   industrial processes. Various components of lipase regulation cascade is
CC   useful in expression methods and systems designed for the production of
CC   heterologous proteins. The present sequence is a DNA encoding P.
CC   alcaligenes LipR, a DNA binding regulator protein
XX
SQ   Sequence 1416 BP; 258 A; 510 C; 453 G; 195 T; 0 U; 0 Other;

  Query Match            36.4%;  Score 327.6;  DB 6;  Length 1416;
  Best Local Similarity  61.0%;  Pred. No. 3.8e-35;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62

Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122

Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164

Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224

Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284

Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344

Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404

Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464

Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524

Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584

Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644

Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704

Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764

Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824

Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543
```

```
Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              ||||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db        482 CGGATCTTGCCGTAAA 467


RESULT 7
AAV23493/c
ID   AAV23493 standard; DNA; 4377 BP.
XX
AC   AAV23493;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Pseudomonas LipQ, LipR, OrfV operon.
XX
KW   Kinase; LipQ; LipR; lipase expression regulator; DNA binding regulator;
KW   sigma 54 promoter; secretion factor; lux-box binding element;
KW   orfV-box binding element; regulation cascade; ss.
XX
OS   Pseudomonas alcaligenes.
XX
FH   Key             Location/Qualifiers
FT   CDS             348. .1377
FT                   /*tag=  a
FT                   /product= "LipQ"
FT   CDS             1443. .2874
FT                   /*tag=  b
FT                   /product= "LipR"
FT   CDS             3531. .4377
FT                   /*tag=  c
FT                   /product= "OrfZ"
FT                   /note= "no stop codon given"
XX
PN   WO9806836-A2.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US014450.
XX
PR   16-AUG-1996;   96US-00699092.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 1998-159528/14.
DR   P-PSDB; AAW53824, AAW53825, AAW53840.
XX
PT   Nucleic acids encoding proteins involved in the lipase regulation cascade
PT   from P. alcaligenes - useful for controlling production and secretion of
PT   heterologous proteins in P. alcaligenes.
XX
PS   Disclosure; Page 42-43; 106pp; English.
XX
CC   This sequence encodes the OrfZ secretion factor, LipQ and lipR kinases of
CC   Pseudomonas alcaligenes. The DNA represents a nucleic acid of the
CC   invention. The nucleic acids encode: (a) a kinase from a Pseudomonad that
CC   regulates the expression of a lipase; (b) a DNA binding regulator from a
CC   Pseudomonad that regulates the expression of a lipase; (c) a Pseudomonas
CC   alcaligenes upstream activating sequence; (d) a P. alcaligenes sigma 54
CC   promoter that regulates the expression of a lipase; (e) a P. alcaligenes
CC   secretion factor selected from XcpP, Q, R, S, T, U, V, W, X, Y and Z, and
CC   OrfV, X, Y; (f) a P. alcaligenes lux-box binding element; and (g) a orfV-
CC   box binding element. The nucleic acids represent parts of a regulation
CC   cascade, comprising at its heart a kinase and a DNA binding regulator.
CC   These sequences can be used for the production of heterologous proteins
CC   in a host cell. The cascade also comprises secretion factors which can
CC   enhance the secretion of produced proteins
XX

SQ   Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;

  Query Match           36.4%;  Score 327.6;  DB 2;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 3e-35;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| ||||||  |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                ||      | ||  | |   |   || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              || | ||  |     |       | ||  | |||||  || ||  |  |          ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| ||||| |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||  ||       ||| |||      || |||   |  |     ||   | |
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||   | |   | | ||    || | |   |||||         ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||    |  ||  |||| ||| | || ||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||   ||       ||| || |  ||| || ||||| || || ||||||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   || ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  |||||||||||  |||||  ||| || |||||||||||||| |||||||||  | |||
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  ||| || || ||  | |||   ||||| | |   ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  |||||||  | |  ||||||| |||||||| | |
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||   ||      ||  |    ||  ||||| ||| ||||| | | |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909


RESULT 8
AAA13904/c
```

```
ID   AAA13904 standard; DNA; 4377 BP.
XX
AC   AAA13904;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Pseudomonas alcaligenes nucleotide sequence SEQ ID NO:28.
XX
KW   Pseudomonas alcaligenes; expression; lipase regulation cascade; kinase;
KW   DNA binding regulator; polymerase; promoter; secretion factor; XcpP;
KW   XcpQ; XcpR; XcpS; XcpT; XcpU; XcpV; XcpW; XcpX; XcpY; XcpZ; OrfV; OrfX;
KW   OrfY; OrfZ; LipQ; LipR; upstream activating sequence; detergent;
KW   cleaning formulation; ss.
XX
OS   Pseudomonas alcaligenes.
XX
PN   US6048710-A.
XX
PD   11-APR-2000.
XX
PF   15-AUG-1997;   97US-00911853.
XX
PR   16-AUG-1996;   96US-00699092.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 2000-316896/27.
XX
PT   Expression vector for producing heterologous proteins in host cells
PT   comprises a nucleic acid encoding a kinase and a DNA binding regulator
PT   which hybridizes under stringent conditions to nucleic acid.
XX
PS   Example 3; Fig 4; 133pp; English.
XX
CC   The present invention describes an expression vector comprising a nucleic
CC   acid encoding a kinase and a DNA binding regulator which hybridises under
CC   stringent conditions to a nucleic acid isolated from Pseudomonas
CC   alcaligenes. Also described are: (1) an isolated plasmid comprising the
CC   above expression vector; (2) a method of transforming a host cell
CC   comprising adding the above plasmid to host cells under appropriate
CC   conditions; (3) a transformed host cell comprising the above expression
CC   vector; and (4) a method for producing a protein comprising the steps of
CC   obtaining a host cell comprising the above expression vector and further
CC   comprising nucleic acid encoding the protein, and culturing the host cell
CC   under conditions for the expression of protein. The expression vector of
CC   the present invention can be used for producing heterologous proteins in
CC   host cells, particularly, lipase in Pseudomonas. Lipases produced can be
CC   used in detergents and cleaning formulations in industrial processes. The
CC   invention provides a higher production level and efficiently express a
CC   heterologous protein. The present sequence represents a Pseudomonas
CC   alcaligenes nucleotide sequence which contains the LipQ, LipR and OrfZ
CC   sequences, from the present invention
XX
SQ   Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;

  Query Match             36.4%;  Score 327.6;  DB 3;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 3e-35;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| |||||| |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                 ||    | ||  | |   |   || |    ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |     |     | ||  |  |||||  || ||  |  |         ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||      || |||   |  |      ||   | |
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||   | |   | | ||    || |  |    ||||||         ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  ||||| |||||||||  ||   | |  ||  |||| ||| | || |||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||    ||          ||| || |  ||| || ||||| || || ||||||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |  ||  ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  |||||||||||  ||||| ||| || ||||||||||||||| |||||||||  | |||
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  ||| || || ||  | |||   ||||| | |   ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||  |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||    ||    ||  |   ||   ||||| ||| ||||| | |  |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909


RESULT 9
AAF30869/c
ID   AAF30869 standard; DNA; 4377 BP.
XX
AC   AAF30869;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Pseudomonas alcaligenes LipQ, LipR and OrfZ gene region.
XX
KW   LipQ; kinase; LipQ; DNA binding regulator; OrfZ; polymerase; lipase; ds.
XX
OS   Pseudomonas alcaligenes.
XX
FH   Key             Location/Qualifiers
FT   protein_bind    257. .275
FT                   /*tag=  a
FT                   /standard_name= "lux-box"
FT   CDS             349. .1377
FT                   /*tag=  b
```

```
FT                   /gene= "lipQ"
FT                   /product= "kinase"
FT                   /note= "encodes AAB82253; this coding region is
FT                   specifically claimed in Claim 1"
FT   CDS             1443. .2858
FT                   /*tag= c
FT                   /gene= "lipR"
FT                   /product= "DNA binding regulator"
FT                   /note= "encodes AAB2254"
FT   stem_loop       2930. .2966
FT                   /*tag= d
FT                   /function= "terminator"
FT   protein_bind    3055. .3067
FT                   /*tag= e
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3059. .3070
FT                   /*tag= f
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3085. .3097
FT                   /*tag= g
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3115. .3147
FT                   /*tag= h
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3118. .3130
FT                   /*tag= i
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3210. .3213
FT                   /*tag= j
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3327. .3339
FT                   /*tag= k
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3349. .3361
FT                   /*tag= l
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3352. .3364
FT                   /*tag= m
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   CDS             3531. .4376
FT                   /*tag= n
FT                   /partial
FT                   /gene= "orfZ"
FT                   /product= "polymerase"
FT                   /note= "encodes AAB82255"
XX
PN   US6225106-B1.
XX
PD   01-MAY-2001.
XX
PF   06-JAN-2000; 2000US-00479409.
XX
PR   16-AUG-1996;   96US-00699092.
PR   15-AUG-1997;   97US-00911853.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 2001-315684/33.
DR   P-PSDB; AAB82253, AAB82254, AAB82255.
XX
PT   Novel isolated nucleic acid encoding kinase from Pseudomonas that can
PT   regulate expression of lipase, useful in expression systems for
PT   production of lipase which is useful in detergents and cleaning
PT   formulations.
XX
PS   Claim 1; Fig 4A-G; 131pp; English.
XX
CC   The present sequence is derived from overlapping sequences of cosmids 71,
CC   201, 505 and 726 that include open reading frames of lipQ (claimed), lipR
CC   and orfZ from Pseudomonas alcaligenes. The cosmids were isolated
CC   following construction of a cosmid library from P. alcaligenes DNA in
CC   Escherichia coli, transformation of the cosmid library into P.
CC   alcaligenes overexpressing lipase, and selection of cosmids stimulating
CC   lipase expression. The lipQ gene, identified by homology to known
CC   sequences, encodes a kinase (see AAB72253) that regulates lipase
CC   expression. The lipR gene encodes a DNA binding regulator (see AAB82254).
CC   The incomplete orfZ gene encodes a polymerase. Another isolated cosmid
CC   (see AAF30870) includes genes encoding P. alcaligenes secretion factors.
CC   A new expression system is provided which comprises components of a
CC   lipase regulation cascade including the kinase, DNA binding regulator,
CC   polymerase, a promoter, an upstream activating sequence, and the
CC   secretion factors. Plasmids and transformed cells are provided, and also
CC   host cells which further comprise a nucleic acid encoding a desired
CC   protein, especially an esterase, hydrolase, lipase, isomerase, mutase,
CC   transferase, kinase or phosphatase (claimed). A hyperproducing strain can
CC   be developed
XX
SQ   Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;

  Query Match           36.4%;  Score 327.6;  DB 4;  Length 4377;
  Best Local Similarity  61.0%;  Pred. No. 3e-35;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||||   ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| ||||||  |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                ||     | ||  | |   |   || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |    |       | ||  | |||||  || ||  |  |        ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| ||||| |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||       || |||   |  |      ||   ||
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||  | |  | | ||   || |  |   ||||||      ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
                |||||  |||||  |||||||||  ||   | |  ||  |||| ||| | || |||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
                |||||    ||        ||| || |  ||| || ||||| || |||  ||||||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   || ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  |||||||||||  ||||| ||| || ||||||||||||||| ||||||||| | |||
```

```
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| ||| ||| || || ||  | |||   |||||  | |   ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||     | | |||   ||      ||  |    ||   ||||| ||| ||||| | | |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909
```

```
RESULT 10
AAD22881/c
ID   AAD22881 standard; DNA; 4377 BP.
XX
AC   AAD22881;
XX
DT   26-FEB-2002  (first entry)
XX
DE   P. alcaligenes DNA comprising cosmids #71, #201, #505 and #726.
XX
KW   DNA binding regulator; lipase regulation cascade; kinase; polymerase;
KW   promoter; upstream activating factor; secretion factor; detergent;
KW   cleaning formulation; industrial process; LipQ; LipR; OrfZ; ds.
XX
OS   Pseudomonas alcaligenes.
XX
FH   Key             Location/Qualifiers
FT   misc_binding    257. .275
FT                   /*tag=  a
FT                   /bound_moiety= "Lux-box (auto-inducer)"
FT   CDS             349. .1377
FT                   /*tag=  b
FT                   /product= "LipQ protein"
FT   CDS             1443. .2858
FT                   /*tag=  c
FT                   /product= "LipR protein"
FT   terminator      2930. .2942
FT                   /*tag=  d
FT   terminator      2954. .2966
FT                   /*tag=  e
FT   misc_binding    3055. .3067
FT                   /*tag=  f
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3058. .3070
FT                   /*tag=  g
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3085. .3097
FT                   /*tag=  h
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3115. .3127
FT                   /*tag=  i
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3118. .3130
FT                   /*tag=  j
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3200. .3212
FT                   /*tag=  k
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3327. .3339
FT                   /*tag=  l
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3349. .3361
FT                   /*tag=  m
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3351. .3364
FT                   /*tag=  n
FT                   /bound_moiety= "OrfV regulator"
FT   CDS             3531. .4376
FT                   /*tag=  o
FT                   /product= "OrfZ protein"
FT                   /note= "CDS does not include stop codon"
FT                   /partial
XX
PN   US6313283-B1.
XX
PD   06-NOV-2001.
XX
PF   06-JAN-2000; 2000US-00479453.
XX
PR   16-AUG-1996;   96US-00699092.
PR   15-AUG-1997;   97US-00911853.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 2002-024912/03.
DR   P-PSDB; AAE13654, AAE13655, AAE13670.
XX
PT   Polynucleotide encoding DNA binding regulator, useful for regulating
PT   expression of lipase especially in pseudomonad, and lipase useful in
PT   detergents and other cleaning formulations and in various industrial
PT   processes.
XX
PS   Example 3; Fig 4; 132pp; English.
XX
CC   The present invention relates to a polynucleotide encoding a DNA binding
CC   regulator that can regulate the expression of a lipase. The invention
CC   also relates to an expression system comprising components of Pseudomonas
CC   alcaligenes lipase regulation cascade which includes kinases, DNA binding
CC   regulators, polymerases, promoters, upstream activating factors and
CC   secretion factors. DNA binding regulators of the invention are useful for
CC   regulating the expression of a lipase, where the lipase is useful in
CC   detergents and other cleaning formulations as well as a number of
CC   industrial processes. Various components of lipase regulation cascade is
CC   useful in expression methods and systems designed for the production of
CC   heterologous proteins. The present sequence is P. alcaligenes DNA
CC   comprising the cosmids #71, #201, #505 and #726 that includes the open
CC   reading frames of LipQ, LipR and a part of OrfZ proteins
XX
SQ   Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;
```

Query Match 36.4%; Score 327.6; DB 6; Length 4377;
Best Local Similarity 61.0%; Pred. No. 3e-35;
Matches 559; Conservative 0; Mismatches 339; Indels 18; Gaps 1;

```
Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  |||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| ||||||  |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                ||      | ||  | |   |   || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
               ||  | ||  |    |       | ||  | |||||  || ||  |  |        ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
```

```
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344

Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404

Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464

Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524

Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584

Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644

Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704

Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764

Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824

Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884

Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900

Db       1924 CGGATCTTGCCGTAAA 1909
```

RESULT 11
ABD16927
ID ABD16927 standard; DNA; 1386 BP.
XX
AC ABD16927;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #15531.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO83356.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 15531; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1386 BP; 218 A; 504 C; 460 G; 204 T; 0 U; 0 Other;

Query Match 30.5%; Score 274.8; DB 11; Length 1386;
Best Local Similarity 60.9%; Pred. No. 3.7e-28;
Matches 447; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

```
Qy        167 GGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGC 226

Db        211 GGACAACGGCGGGCCGTCGCCGCGCGCGCATTCCAGGCCGAGGTCCTGGGCCTCGATCTG 270

Qy        227 GCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCC 286

Db        271 CTTTCCCTCTGCCAGGACCAACCCGCGACGTACCCGGTTGGCCAGTTCGCGGACATTGCC 330

Qy        287 CGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTG 346

Db        331 CGGCCAGGCGTGCATCGCCATCGCCGCCAGCGCGTCCTCGCTGAAGCGGCGCGGGCGGCG 390

Qy        347 CCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGG 406

Db        391 GCCGGTCTCCAGGCTGTAGAGCCGGGAGAAATGCTGGGCCAGCAGCGGCAGGTCCTCGCT 450

Qy        407 CCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCT 466

Db        451 CCGTTCGCGCAACGGCGCGGTGCCGACCTGCAGGACGTTGAGCCGGTAGTAGAGGTCCTC 510

Qy        467 CCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCG 526

Db        511 GCGGAACGTCCCGGCATCGATCGCCGCCTCCAGGTCGACATGGGTCGCCGCCAGCACCCG 570

Qy        527 CACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAA 586

Db        571 CACGTCCACCTCGACCGGCCGGCTACCGCCGACCCGCTCGATCTGCTGTTCCTGGAGAAA 630

Qy        587 ACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGT 646

Db        631 GCGCAACAGGTTGGCCTGCAACTCCAGCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGT 690

Qy        647 CCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCC 706

Db        691 GCCGCCGTCGGCCGCCTCGATCCGCCCGACCTTGCGCTGGTGCGCGCCGGTGAAGGCGCC 750

Qy        707 GGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCAC 766
```

```
Db        751 CTTCTCGTGGCCGAACAGCTCGGACTGGATCAGGTGCTCGGGAATCGCCCCGCAGTTGAT 810

Qy        767 GAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTC 826
              |||||||||| | |   ||| |   | |||||||| |  ||||| | |   ||
Db        811 CGCGACGAACGGCTTGTCCGCGCGCCGCGACTGCCGGTGCAGGGTACGCGCGACCAGTTC 870

Qy        827 CTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCG 886
              |||||| ||||| | ||||||    | ||| | || ||  ||| |||||||||| |||
Db        871 CTTGCCGGTGCCGCTCTCGCCACGGATCAGTACCGGCGAGTCGGTCGGCGCCAGCTTCGC 930

Qy        887 CACCTGCGCGAACA 900
              || | ||  |  ||
Db        931 CAGCAGCCGGCGCA 944


RESULT 12
ABD16704/c
ID   ABD16704 standard; DNA; 1461 BP.
XX
AC   ABD16704;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #15308.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO83133.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 15308; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 1461 BP; 232 A; 474 C; 513 G; 242 T; 0 U; 0 Other;

  Query Match           30.5%;  Score 274.8;  DB 11;  Length 1461;
  Best Local Similarity 60.9%;  Pred. No. 3.7e-28;

  Matches  447;  Conservative    0;  Mismatches  287;  Indels    0;  Gaps    0;

Qy        167 GGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGC 226
              ||| |  |||||         |   ||| |   || || |||||||       |||||||
Db       1311 GGACAACGGCGGGCCGTCGCCGCGCGCGCATTCCAGGCCGAGGTCCTGGGCCTCGATCTG 1252

Qy        227 GCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCC 286
              | ||| | || || ||||   ||||  | ||   |||  |||| || ||||| |||||
Db       1251 CTTTCCCTCTGCCAGGACCAACCCGCGACGTACCCGGTTGGCCAGTTCGCGGACATTGCC 1192

Qy        287 CGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTG 346
              |||||| |    | ||| || |||||  |||| ||| |    |   | ||||     |
Db       1191 CGGCCAGGCGTGCATCGCCATCGCCGCCAGCGCGTCCTCGCTGAAGCGGCGCGGGCGGCG 1132

Qy        347 CCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGG 406
              ||  || || |  |     ||       |   || ||| |   ||  |  |||| ||
Db       1131 GCCGGTCTCCAGGCTGTAGAGCCGGGAGAAATGCTGGGCCAGCAGCGGCAGGTCCTCGCT 1072

Qy        407 CCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCT 466
              ||| || |||| |||||     | || || ||| ||    ||||| | || ||||
Db       1071 CCGTTCGCGCAACGGCGCGGTGCCGACCTGCAGGACGTTGAGCCGGTAGTAGAGGTCCTC 1012

Qy        467 CCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCG 526
              |||||  |||| |   | |  |||| | || ||    || |||||||| | || |||
Db       1011 GCGGAACGTCCCGGCATCGATCGCCGCCTCCAGGTCGACATGGGTCGCCGCCAGCACCCG 952

Qy        527 CACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAA 586
              ||||||| ||      |  ||  |    || ||||||   |  |       ||| | |||
Db        951 CACGTCCACCTCGACCGGCCGGCTACCGCCGACCCGCTCGATCTGCTGTTCCTGGAGAAA 892

Qy        587 ACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGT 646
              ||||  ||||    |||||| || || |||||| ||||||| ||||||||  | ||| ||
Db        891 GCGCAACAGGTTGGCCTGCAACTCCAGCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGT 832

Qy        647 CCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCC 706
              |||||  || |||||||||  | ||||  ||      |  | |||||||  | |   |||||
Db        831 GCCGCCGTCGGCCGCCTCGATCCGCCCGACCTTGCGCTGGTGCGCGCCGGTGAAGGCGCC 772

Qy        707 GGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCAC 766
              |||||  |||||||||||  || |||| |    ||||| |||||| |||| |  |
Db        771 CTTCTCGTGGCCGAACAGCTCGGACTGGATCAGGTGCTCGGGAATCGCCCCGCAGTTGAT 712

Qy        767 GAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTC 826
              |||||||||| | |   ||| |   | |||||||| |  ||||| | |   ||
Db        711 CGCGACGAACGGCTTGTCCGCGCGCCGCGACTGCCGGTGCAGGGTACGCGCGACCAGTTC 652

Qy        827 CTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCG 886
              |||||| ||||| | ||||||    | ||| | || ||  ||| |||||||||| |||
Db        651 CTTGCCGGTGCCGCTCTCGCCACGGATCAGTACCGGCGAGTCGGTCGGCGCCAGCTTCGC 592

Qy        887 CACCTGCGCGAACA 900
              || | ||  |  ||
Db        591 CAGCAGCCGGCGCA 578


RESULT 13
ACL72552/c
ID   ACL72552 standard; DNA; 972 BP.
XX
AC   ACL72552;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus gene sequence, seq id 9015.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression;
KW   gene; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
```

```
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 9015; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 972 BP; 146 A; 345 C; 348 G; 133 T; 0 U; 0 Other;

  Query Match             29.8%;  Score 267.8;  DB 14;  Length 972;
  Best Local Similarity   60.0%;  Pred. No. 3.3e-27;
  Matches  465;  Conservative    0;  Mismatches  307;  Indels    3;  Gaps    1;

Qy        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
              | || || |   | | |  | ||| | ||| |     |   ||||  | |  |||||
Db        814 CATCCAGGGGCGCAGGCGCAGTCGGCGCGTTGACACCCTGAGCGGCTGCGGGAGGCGGCA 755

Qy        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240
              | |    | |    | ||| ||  | |  ||  |  ||||        ||  |  ||
Db        754 GGGCGGCCCGGTCCACATCGGGCAGGGGCAGGTGCTCTTCGAGAATCTCCCCTTCACAGA 695

Qy        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300
              ||||||||||||  | |    |||||||||||||||| ||||| || |||||  |
Db        694 GCACCACGGCGCTCTCGATACAGTTCTCCAGCTCCCGCACGTTTCCGGGCCAGCGGTAGC 635

Qy        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360
              |  ||  |  | |   |||  | ||    ||   | ||||||  | |||     |  |||  |
Db        634 GCTTGAGGCGCTCCACCGCGGCGGCGCTGAGGCGGGGCGGCGTCAGCCGGTGCCTCCGGG 575

Qy        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGC---ACCGCGTCCCCGGGCCGCTCCCGCA 417
                  |  |          |||||  |||   |   | ||| ||   ||||||||||
Db        574 CGACGGCGGCGACGAAGTGGCGGGCGAGCCGCTCGATGTCCTCCGCGCCGCGCTCCCGCA 515

Qy        418 GCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCC 477
              ||||||||||||   | ||| | ||||   |  ||| | || |||||    |||||||  |
Db        514 GCGGCGGCAGCACCACCTCGACCACCTTGATGCGGTAGTAGAGGTCCTCGCGGAAGCGGC 455

Qy        478 CCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCC 537
              |||||  ||||||      | || ||||| || || |||||||   |  |||||||| ||
Db        454 CCTCGGCCACCATGCGGGCCAGGTCCCGATGGGTGGCCGCGACGATGCGCACGTCCACCT 395

Qy        538 GCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGC 597
              |  |||  | ||    ||||| |||    |  ||||    ||| |  |  |||||||
Db        394 TCACGGCCTGGGTGCCTCCCACGCGCTCGAACTCGCGATCCTGGATGACCCGCAGCAACT 335

Qy        598 GCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCG 657
              |||||||||  ||  |||||||||||| | ||||||||  | ||  ||| |||||    |
Db        334 TGCCCTGCACCGGCAGGGGCAGCTCGCCAATCTCGTCGATGAACACCGTGCCGCCGCTGG 275

Qy        658 CCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACC 717
              ||||||||| ||   || | |        | ||||||  | |   ||||||  |||||  |
Db        274 CCGCCTCGAACTTGCCGGGCACGCGGTGGTCCGCGCCGGTGAAGGCGCCGCGTTCGTGGC 215

Qy        718 CGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACG 777
              |||| ||||||  ||||||  |||   |||| | |||||||||  | |||    | ||| |
Db        214 CGAAGAGCTCGTTCTCGATGAGCGTGGCGGGCAGCGCCGCGCAGTCCACCTTGATGAAGG 155

Qy        778 GCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGC 837
              |||| | ||  || || |     |  ||| || || || || | |  |||||||||  |||
Db        154 GCTGGTCCCTGCGGGGACCATTCACGTGGACGGCACGGGCGAACAGCTCCTTGCCGCTGC 95

Qy        838 CCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTG 892
              | | ||||||   || ||||| |||||| ||| | |||||  ||||||||||| |
Db         94 CACTCTCGCCGCGCAGCAGCACCGTCGCATCGGTGGGCGCGGCCTTGCGCACCAG 40


RESULT 14
ACL69201/c
ID   ACL69201 standard; DNA; 1053 BP.
XX
AC   ACL69201;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus gene sequence, seq id 5664.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression;
KW   gene; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 5664; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 1053 BP; 147 A; 333 C; 428 G; 145 T; 0 U; 0 Other;

  Query Match             29.8%;  Score 267.8;  DB 14;  Length 1053;
  Best Local Similarity   61.3%;  Pred. No. 3.3e-27;
  Matches  431;  Conservative    0;  Mismatches  272;  Indels    0;  Gaps    0;

Qy        198 TCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGA 257
              || ||||  |||||      ||| | | |    ||  | ||||| |   ||||  || |
Db        848 TCCAGCCACAGGTGCCGAGGCTCCACCACCGGCCCATCCGCGAGGATGGCGGCCCGCTCG 789
```

```
Qy        258 ACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGC 317
              |  || ||   |||||| ||||||||||||||||| |    |||| | |    |||
Db        788 AGCGCATTGCGCAGCTCGCGGACGTTGCCCGGCCAGGTGAACGCCTCCAGCCGCGCCACG 729

Qy        318 GCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGC 377
              |||||||  ||||     |      |   | |||| | |           ||   |||||||
Db        728 GCCTCCGGTGACACGCGCAAGCCTGGACGACCCAGCTCGTCTCCGATGCGGCGCAGCAGC 669

Qy        378 GACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCC 437
              |||  |  |   ||  |  |||| || ||||||| |||||  | ||||||||||||
Db        668 AGCTCCGACAACGGCCGCAGGTCCTCGCGCCGCTCGCGCAGGGACGGCAGCCGGATGGGG 609

Qy        438 AGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCG 497
              | |||   |||||||   || |||||    |||||    ||       || |||||||  |
Db        608 AACACGGCCAGCCGGTGGTAGAGGTCCTCGCGGAACTCACCGCGCGCCATCATCGCCTTG 549

Qy        498 AGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCC 557
              |||||  ||||  || |||||  | |   |||||||| ||    | |||  |||||    |||
Db        548 AGATCGCGGTTGGTGGCCGCCACCCAGCGCACGTCGGCTTCCAGGGTCCGCGTACCGCCC 489

Qy        558 ACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGC 617
              || ||    |   ||    || |  |  ||||||||     |||| | ||   | |
Db        488 ACGCGTTCGAAGCGGCGTTCTTGGAGGACGCGCAGCAGCTTGGCCTGGAGCTCGGCCTTC 429

Qy        618 AGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCC 677
              ||||||||||||||||  |   |  || || |||| ||||  ||||    | ||   |
Db        428 AGCTCGCCGACCTCGTCCAAGAAGAAGGTGCCTCCCTGCGCCAGCTCGATGCGGCCGCGC 369

Qy        678 CGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATC 737
              |||   || | ||||||  | |  ||||||   |||||  ||||| |||||| ||| |
Db        368 CGCTGGGCCACCGCGCCGGTGAACGCGCCCTTCTCGTGTCCGAAGAGCTCGCTCTCCAGG 309

Qy        738 TGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTC 797
              |||  || |  | |||||||||  |  |||  ||||||  |||    | |   |||  |||
Db        308 AGCGTCTCCGTCAGCGCCGCGCAGTTAACGGCCACGAAAGGCCCGTCCGCGCGCTCACTC 249

Qy        798 AGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGC 857
              | ||||||  || |||||   |  |||||||| || || | |||  ||||||| ||||
Db        248 CACTGGTGCAGGGCCCGCGCGGTCACCTCCTTGCCGGTCCCGCTCTCCCCCACCAGCAGC 189

Qy        858 AGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              | ||| ||||   | |||| ||||||  |||| |    | | ||
Db        188 ACCGTGGCCTGCGTGGGCGACACCTTCCGCAGCGCTTCCACCA 146


RESULT 15
ACL64578
ID   ACL64578 standard; DNA; 11812 BP.
XX
AC   ACL64578;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus DNA fragment, seq id 1041.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 1; SEQ ID NO 1041; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC   set of 1849 contig and singleton sequences comprising coding sequences,
CC   DNA replication elements, promoters and other regulatory elements from
CC   the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO
XX
SQ   Sequence 11812 BP; 1923 A; 4251 C; 3876 G; 1761 T; 0 U; 1 Other;

  Query Match             29.8%;  Score 267.8;  DB 14;  Length 11812;
  Best Local Similarity   61.3%;  Pred. No. 2e-27;
  Matches  431;  Conservative    0;  Mismatches  272;  Indels    0;  Gaps    0;

Qy        198 TCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGA 257
              || ||||  |||||      ||| | | |    ||  | |||||  |   |||| || |
Db      10964 TCCAGCCACAGGTGCCGAGGCTCCACCACCGGCCCATCCGCGAGGATGGCGGCCCGCTCG 11023

Qy        258 ACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGC 317
              |  || ||   |||||| ||||||||||||||||| |    |||| | |    |||
Db      11024 AGCGCATTGCGCAGCTCGCGGACGTTGCCCGGCCAGGTGAACGCCTCCAGCCGCGCCACG 11083

Qy        318 GCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGC 377
              |||||||  ||||     |      |   | |||| | |           ||   |||||||
Db      11084 GCCTCCGGTGACACGCGCAAGCCTGGACGACCCAGCTCGTCTCCGATGCGGCGCAGCAGC 11143

Qy        378 GACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCC 437
              |||  |  |   ||  |  |||| || ||||||| |||||  | ||||||||||||
Db      11144 AGCTCCGACAACGGCCGCAGGTCCTCGCGCCGCTCGCGCAGGGACGGCAGCCGGATGGGG 11203

Qy        438 AGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCG 497
              | |||   |||||||   || |||||    |||||    ||       || |||||||  |
Db      11204 AACACGGCCAGCCGGTGGTAGAGGTCCTCGCGGAACTCACCGCGCGCCATCATCGCCTTG 11263

Qy        498 AGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCC 557
              |||||  ||||  || |||||  | |   |||| ||||  ||    | |||  |||||    |||
Db      11264 AGATCGCGGTTGGTGGCCGCCACCCAGCGCACGTCGGCTTCCAGGGTCCGCGTACCGCCC 11323

Qy        558 ACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGC 617
              || ||    |   ||    || |  |  ||||||||     |||| | ||   | |
Db      11324 ACGCGTTCGAAGCGGCGTTCTTGGAGGACGCGCAGCAGCTTGGCCTGGAGCTCGGCCTTC 11383

Qy        618 AGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCC 677
              ||||||||||||||||  |   |  || || |||| ||||  ||||    | ||   |
Db      11384 AGCTCGCCGACCTCGTCCAAGAAGAAGGTGCCTCCCTGCGCCAGCTCGATGCGGCCGCGC 11443

Qy        678 CGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATC 737
              |||   || | ||||||  | |  ||||||   |||||  ||||| |||||| ||| |
Db      11444 CGCTGGGCCACCGCGCCGGTGAACGCGCCCTTCTCGTGTCCGAAGAGCTCGCTCTCCAGG 11503

Qy        738 TGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTC 797
              |||  || |  | |||||||||  |  |||  ||||||  |||    | |   |||  |||
Db      11504 AGCGTCTCCGTCAGCGCCGCGCAGTTAACGGCCACGAAAGGCCCGTCCGCGCGCTCACTC 11563

Qy        798 AGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGC 857
              | ||||||  || |||||   |  |||||||| || || | |||  ||||||| ||||
Db      11564 CACTGGTGCAGGGCCCGCGCGGTCACCTCCTTGCCGGTCCCGCTCTCCCCCACCAGCAGC 11623

Qy        858 AGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              | ||| ||||   | |||| ||||||  |||| |    | | ||
```

Db 11624 ACCGTGGCCTGCGTGGGCGACACCTTCCGCAGCGCTTCCACCA 11666

Search completed: April 27, 2006, 07:59:50
Job time : 438.991 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 185.684 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_1_900
Perfect score: 900
Sequence: 1 tgggtatcccgagccgctgg.....cttgcgcacctgcgcgaaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		336	37.3	1077	3	US-09-252-991A-10719	Sequence 10719, A
c 2		336	37.3	1173	3	US-09-252-991A-10448	Sequence 10448, A
c 3		327.6	36.4	1416	3	US-08-911-853-3	Sequence 3, Appli
c 4		327.6	36.4	1416	3	US-09-479-409-3	Sequence 3, Appli
c 5		327.6	36.4	1416	3	US-09-479-453-3	Sequence 3, Appli
c 6		327.6	36.4	4377	3	US-08-911-853-28	Sequence 28, Appl
c 7		327.6	36.4	4377	3	US-09-479-409-28	Sequence 28, Appl
c 8		327.6	36.4	4377	3	US-09-479-453-28	Sequence 28, Appl
9		274.8	30.5	1386	3	US-09-252-991A-15531	Sequence 15531, A
c 10		274.8	30.5	1461	3	US-09-252-991A-15308	Sequence 15308, A
c 11		267.8	29.8	972	3	US-09-902-540-9015	Sequence 9015, Ap
c 12		267.8	29.8	1053	3	US-09-902-540-5664	Sequence 5664, Ap
13		267.8	29.8	11812	3	US-09-902-540-1041	Sequence 1041, Ap
14		267.8	29.8	12849	3	US-09-902-540-963	Sequence 963, App
c 15		266	29.6	2379	3	US-09-252-991A-9205	Sequence 9205, Ap
c 16		264.6	29.4	1617	3	US-09-902-540-5384	Sequence 5384, Ap
17		264.6	29.4	19053	3	US-09-710-262E-20	Sequence 20, Appl
c 18		264.6	29.4	34552	3	US-09-902-540-1262	Sequence 1262, Ap
c 19		260.8	29.0	837	3	US-09-252-991A-9123	Sequence 9123, Ap
c 20		260	28.9	1362	3	US-09-902-540-5154	Sequence 5154, Ap
c 21		260	28.9	29559	3	US-09-902-540-1254	Sequence 1254, Ap
c 22		256.2	28.5	1344	3	US-09-902-540-7558	Sequence 7558, Ap
23		256.2	28.5	6155	3	US-09-902-540-737	Sequence 737, App
c 24		254.6	28.3	3699	3	US-09-902-540-3731	Sequence 3731, Ap
25		254.6	28.3	19112	3	US-09-902-540-1181	Sequence 1181, Ap
c 26		250.4	27.8	1107	3	US-09-902-540-8250	Sequence 8250, Ap
27		250.4	27.8	6698	3	US-09-902-540-852	Sequence 852, App
c 28		248.8	27.6	1356	3	US-09-902-540-8909	Sequence 8909, Ap
29		248.8	27.6	9091	3	US-09-902-540-949	Sequence 949, App
c 30		247.8	27.5	1350	3	US-09-902-540-4545	Sequence 4545, Ap
c 31		247.8	27.5	10835	3	US-09-902-540-1031	Sequence 1031, Ap
c 32		245.6	27.3	1422	3	US-09-902-540-3119	Sequence 3119, Ap
c 33		245.6	27.3	15268	3	US-09-902-540-1142	Sequence 1142, Ap
c 34		245.2	27.2	1341	3	US-09-902-540-8826	Sequence 8826, Ap
c 35		245.2	27.2	8765	3	US-09-902-540-939	Sequence 939, App
c 36		244.8	27.2	1434	3	US-09-902-540-5210	Sequence 5210, Ap
c 37		244.8	27.2	31826	3	US-09-902-540-1256	Sequence 1256, Ap
c 38		244.2	27.1	1491	3	US-09-902-540-3335	Sequence 3335, Ap
c 39		244.2	27.1	16387	3	US-09-902-540-1156	Sequence 1156, Ap
c 40		243.4	27.0	1252	3	US-09-902-540-6749	Sequence 6749, Ap
41		243.4	27.0	4142	3	US-09-902-540-546	Sequence 546, App
c 42		243	27.0	1161	3	US-09-252-991A-9202	Sequence 9202, Ap
43		243	27.0	1209	3	US-09-252-991A-8857	Sequence 8857, Ap
c 44		241.2	26.8	1358	3	US-09-902-540-3270	Sequence 3270, Ap
c 45		241.2	26.8	17480	3	US-09-902-540-1151	Sequence 1151, Ap

ALIGNMENTS

RESULT 1
US-09-252-991A-10719
; Sequence 10719, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10719
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10719

Query Match 37.3%; Score 336; DB 3; Length 1077;
Best Local Similarity 62.3%; Pred. No. 5.8e-46;
Matches 578; Conservative 0; Mismatches 320; Indels 30; Gaps 2;

```
Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || |||||||| |||||||||| |   |||||||||||||| |||   |||
Db         11 GGAATGCCCAGGCGCTGGCGGCGTTCCCACAGGCATTTGCGGCTGATGCCCAGCTTGCGG 70

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || | |||||||||| |||| | |||||||||||||| || |||||| |  |||| ||| |
Db         71 GCCAACTCGGTCTCGGTCATGTGGTCCTGGTGCTCGAGGACGAAGTGCTGGAAGTAGTCT 130

Qy        123 TCGAGCGA--------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATG 162
              || |||||                   |   |  ||| ||| | |    ||   ||| |
Db        131 TCCAGCGAAAGGTCCTCGGTGGGCTCGTGGTTGCTGGCCGCGGCGCCCGCCGGGCCGAGG 190

Qy        163 CGGAGGAGACGGGCGGAGGCGGCCGCG----------GCGGGTCGTCGAGCCCGAGGTGG 212
              | ||||     |||| || || || ||          ||  |  |||||  ||  ||  ||
Db        191 CAGAGGTCGGCGGCGAAGTCGTCCTCGAGGTCGTCCAGCTCGATGTCGATACCCAGCAGG 250

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272
                |   | |||  |   || | ||  ||| | ||| || || |  | ||||||  |||||
Db        251 TCGGCGTGGATTTCCTGGCTCTCGCTGAGGATCACCGCCCGCTCGATGGCGTTTTCCAGT 310

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332
```

```
Db        311 TCGCGCACGTTGCCCGGCCATGGGTAATGGCGGATCGCCTGTTCGGCGTCGTGGGCGAAG 370

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392

Db        371 CTGAGCGGCGCGCGCCCCATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCC 430

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452

Db        431 ATCACGTCGTTGCCACGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGG 490

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512

Db        491 TAATAGAGGTCTTCGCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTG 550

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572

Db        551 GCGGCGATCAGGCGCACGTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCG 610

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632

Db        611 CCTTCCTGCAATACGCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCG 670

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692

Db        671 TCGAGGAACAGGGTGCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCA 730

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752

Db        731 CCGGTGAAGGCGCCTTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTCCGGAATC 790

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812

Db        791 GCCGCACAGTTCACCGAGATCAGCGGCGCCTTGGCGCGCTTCGACAGGTTGTGCAGGGCG 850

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872

Db        851 CGGGCGACCAGTTCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACCGTGGAGTCGGTC 910

Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900

Db        911 GGCGCGACCTTGCGAATCTTGCTGTACA 938


RESULT 2
US-09-252-991A-10448/c
; Sequence 10448, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10448
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10448

  Query Match             37.3%;  Score 336;  DB 3;  Length 1173;
  Best Local Similarity   62.3%;  Pred. No. 5.8e-46;
  Matches  578;  Conservative    0;  Mismatches  320;  Indels   30;  Gaps    2;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62

Db       1139 GGAATGCCCAGGCGCTGGCGGCGTTCCCACAGGCATTTGCGGCTGATGCCCAGCTTGCGG 1080


Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122

Db       1079 GCCAACTCGGTCTCGGTCATGTGGTCCTGGTGCTCGAGGACGAAGTGCTGGAAGTAGTCT 1020

Qy        123 TCGAGCGA--------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATG 162

Db       1019 TCCAGCGAAAGGTCCTCGGTGGGCTCGTGGTTGCTGGCCGCGGCGCCCGCCGGGCCGAGG 960

Qy        163 CGGAGGAGACGGGCGGAGGCGGCCGCG----------GCGGGTCGTCGAGCCCGAGGTGG 212

Db        959 CAGAGGTCGGCGGCGAAGTCGTCCTCGAGGTCGTCCAGCTCGATGTCGATACCCAGCAGG 900

Qy        213 GCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGC 272

Db        899 TCGGCGTGGATTTCCTGGCTCTCGCTGAGGATCACCGCCCGCTCGATGGCGTTTTCCAGT 840

Qy        273 TCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGC 332

Db        839 TCGCGCACGTTGCCCGGCCATGGGTAATGGCGGATCGCCTGTTCGGCGTCGTGGGCGAAG 780

Qy        333 GCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGC 392

Db        779 CTGAGCGGCGCGCGCCCCATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCC 720

Qy        393 ACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGG 452

Db        719 ATCACGTCGTTGCCACGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGG 660

Qy        453 AAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTC 512

Db        659 TAATAGAGGTCTTCGCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTG 600

Qy        513 GCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCG 572

Db        599 GCGGCGATCAGGCGCACGTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCG 540

Qy        573 CCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCG 632

Db        539 CCTTCCTGCAATACGCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCG 480

Qy        633 TCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCG 692

Db        479 TCGAGGAACAGGGTGCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCA 420

Qy        693 CCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATC 752

Db        419 CCGGTGAAGGCGCCTTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTCCGGAATC 360

Qy        753 GCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCG 812

Db        359 GCCGCACAGTTCACCGAGATCAGCGGCGCCTTGGCGCGCTTCGACAGGTTGTGCAGGGCG 300

Qy        813 CGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTC 872

Db        299 CGGGCGACCAGTTCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACCGTGGAGTCGGTC 240

Qy        873 GGCGCCACCTTGCGCACCTGCGCGAACA 900

Db        239 GGCGCGACCTTGCGAATCTTGCTGTACA 212


RESULT 3
US-08-911-853-3/c
; Sequence 3, Application US/08911853
; Patent No. 6048710
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Genencor International
;       STREET:  925 Page Mill Road
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304-1013
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/911,853
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/699,092
;       FILING DATE:  16-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Glaister, Debra J
;       REGISTRATION NUMBER:  33,888
;       REFERENCE/DOCKET NUMBER:  GC361-2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-846-7620
;       TELEFAX:  650-845-6504
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1416 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-911-853-3

  Query Match             36.4%;  Score 327.6;  DB 3;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 1.3e-44;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| | ||||||||||||||||||    ||
Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| |||||| |  |||| ||| ||
Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                ||    | ||  | |   |  || |    ||
Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |     |      | ||  |  |||||  || ||  |  |        ||
Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |  |||| ||  ||| | ||| ||||| |  | ||| |||||||||| || ||||||
Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||        ||| |||      || |||   |  |    ||   | |
Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||    | |   | | ||    || |  |     |||||          ||||  ||
Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||   | |  ||  ||||  |||  | ||  |||||
Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
                |||||    ||        |||  || |  |||  || |||||  ||  || |||||||
Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   ||  ||| || ||||| || |  ||||||  |||||
Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              | ||||||||||||  ||||| ||| || ||||||||||||||| |||||||| | |||
Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  ||| || || ||  | |||   |||||  |  |   ||
Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||   ||      ||  |    ||   ||||| ||| ||||| | |  |
Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db        482 CGGATCTTGCCGTAAA 467


RESULT 4
US-09-479-409-3/c
; Sequence 3, Application US/09479409
; Patent No. 6225106
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,409
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1416 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-409-3

  Query Match             36.4%;  Score 327.6;  DB 3;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 1.3e-44;
```

```
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||||    ||
Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| |||||| |  |||| ||| ||
Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                  ||      | ||  | |    |   || |      ||
Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |    |       | ||  |  |||||  || ||  |  |         ||
Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||       || |||   |  |      ||   | |
Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||   | |   | | ||    || |  |    ||||||         ||||  ||
Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||   | |  ||  |||| ||| | || |||||
Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||    ||         ||| || |  ||| || ||||| || ||  ||||||||
Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   ||  ||| || ||||| || |  ||||||  |||||
Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  ||||||||||  |||||  ||| || |||||||||||||| |||||||||  | |||
Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || |||||  || |||||  |||  ||| || || ||  | |||   |||||  | |   ||
Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  |||||| | |  |||||||| |||||||| | |
Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||   ||      ||  |    ||  ||||| ||| ||||| | | |
Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||  | ||| |   | |  ||| | || || ||||||
Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db        482 CGGATCTTGCCGTAAA 467


RESULT 5
US-09-479-453-3/c
; Sequence 3, Application US/09479453
; Patent No. 6313283
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,453
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1416 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-453-3

  Query Match             36.4%;  Score 327.6;  DB 3;  Length 1416;
  Best Local Similarity   61.0%;  Pred. No. 1.3e-44;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||||    ||
Db       1382 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 1323

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| |||||| |  |||| ||| ||
Db       1322 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 1263

Qy        123 TCGAGCGA------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                  ||      | ||  | |    |   || |      ||
Db       1262 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 1203

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |    |       | ||  |  |||||  || ||  |  |         ||
Db       1202 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 1143

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       1142 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 1083

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||          ||| |||       || |||   |  |      ||   | |
Db       1082 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 1023

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||   | |   | | ||    || |  |    ||||||         ||||  ||
Db       1022 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 963

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||   | |  ||  |||| ||| | || |||||
Db        962 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 903
```

```
Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||    ||         ||| || |  ||| || ||||| || || ||||||||
Db        902 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 843

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   ||  ||| || ||||| || |  ||||||  |||||
Db        842 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 783

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  ||||||||||  ||||| ||| || ||||||||||||| ||||||||| | |||
Db        782 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 723

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  ||| || || ||  | |||   ||||| | |   ||
Db        722 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 663

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db        662 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 603

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||   ||      ||  |    ||  ||||| ||| ||||| | |  |
Db        602 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 543

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db        542 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 483

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db        482 CGGATCTTGCCGTAAA 467


RESULT 6
US-08-911-853-28/c
; Sequence 28, Application US/08911853
; Patent No. 6048710
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/911,853
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/699,092
;      FILING DATE:  16-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-911-853-28

  Query Match           36.4%;  Score 327.6;  DB 3;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 1.2e-44;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || ||  || ||||| || ||  ||||||| |  |||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| ||||||  |  ||||  ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA-----------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                 ||      | ||  | |   |   || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |    |      | ||  | |||||  || ||  |  |          ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| ||||| |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||      ||| |||      || |||  |  |    ||  | ||
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||  | |  | | ||   || | |    ||||||        ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||   | |  ||  |||| ||| | || |||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||    ||         ||| || |  ||| || ||||| || || ||||||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   ||  ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  ||||||||||  ||||| ||| || ||||||||||||| ||||||||| | |||
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  ||| || || ||  | |||   ||||| | |   ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||   ||      ||  |    ||  ||||| ||| ||||| | |  |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||   | ||| |   | |  ||| | || || ||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909


RESULT 7
```

```
US-09-479-409-28/c
; Sequence 28, Application US/09479409
; Patent No. 6225106
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,409
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-409-28

  Query Match             36.4%;  Score 327.6;  DB 3;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 1.2e-44;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || |||||| |  ||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  |||||||||||||| |||||| |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                  ||      | ||  | |   |   || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |     |       | ||  |  |||||  || ||  |  |       ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |  |||| ||  ||| | ||| |||||  |  | ||| ||||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||       ||| |||      || |||   |  |    ||   | |
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||   | |   | | ||    || |  |    ||||||        ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              ||||| ||||| |||||||||  ||   | |  ||  |||| ||| | || |||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||    ||         ||| || |  ||| || ||||| || || ||||||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | | |   ||  ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  |||||||||||  |||||  ||| || |||||||||||||| |||||||| | |||
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  |||| || || ||  | |||   |||| | |   ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||  |||||  |||||||| |||  ||||||| | |  |||||||| |||||||| | |
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    | | |||  ||    || |    ||  ||||| ||| ||||| | |  |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||   |||||||    | ||| |   | |  ||| | || || ||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              || | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909


RESULT 8
US-09-479-453-28/c
; Sequence 28, Application US/09479453
; Patent No. 6313283
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,453
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
```

```
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-453-28

  Query Match             36.4%;  Score 327.6;  DB 3;  Length 4377;
  Best Local Similarity   61.0%;  Pred. No. 1.2e-44;
  Matches  559;  Conservative    0;  Mismatches  339;  Indels   18;  Gaps    1;

Qy          3 GGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGACGG 62
              || || || || ||||| || || ||||||| |  ||||||||||||||||||    ||
Db       2824 GGAATGCCCAGGCGCTGACGGCGCTCCCACAGGCACTTGCGGCTGATGCCGAGCTTGCGC 2765

Qy         63 GCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGCCC 122
              || |  |||||||| ||||  |  ||||||||||||||| |||||| |  |||| ||| ||
Db       2764 GCCAGTTCGGTCTCGGTCATGTGATCCTGGTGCTCCAGTACGAAGTGCTGGAAGTAGTCC 2705

Qy        123 TCGAGCGA------------------GTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG 164
              || |||||                 ||     | ||  | |   |   || |     ||
Db       2704 TCCAGCGACAGGTCCTCGGTCGGTTCGTGGTTGGCCGCGGTCTGCTGTGGCTGTTCGCCG 2645

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              ||  | ||  |    |     | ||  |  |||||  || ||  |  |        ||
Db       2644 AAGTCGCCGTCCTCCAGGTCGTCCAGCTCGATGTCGATGCCCAGCAGCTCGGCGGAAATT 2585

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |   |||| ||  ||| | ||| |||||  |  | ||| |||||||||| || ||||||
Db       2584 TCCGCGCCCTCGCAGAGGATCACCGCGCGCTCGATGGCATTCTCCAGCTCGCGCACGTTG 2525

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||||| ||       ||| |||     || |||    |  |    ||   | |
Db       2524 CCCGGCCAGGGGTAGTGGCGGATCGCCTGCTCGGCATCCTGAGCGAAGCGCAGGTCCTCG 2465

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | |||||   |  |  | ||    || |  |   ||||||      ||||  ||
Db       2464 CGGCCCATGCGCTGGCACTGGCGGACGAGGAAGGCGCGGGCGATCTCGTTGACGTCGGCG 2405

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              |||||  |||||  |||||||||  ||   | |  ||  |||| ||| | || |||||
Db       2404 CCGCGCTCGCGCAGTGGCGGCAGCTTGAGGGCGATGACGTGCAGGCGGTAGTAGAGGTCC 2345

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||    ||      ||| || |  ||| || ||||| || ||  ||||||||
Db       2344 TCGCGGAACTGGCCGGTCTTGGCCAGCGTCTTGAGGTCGCGGTGGGTAGCGGCGATCAGG 2285

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584
              || || ||  ||  | |   ||  ||| || ||||| || |  ||||||  |||||
Db       2284 CGTACATCGACCTTCTGTGACTGCACCGAGCCGACCCGACGGATCTCGCCCTCCTGCAGC 2225

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644
              |  ||||||||||  |||||  ||| ||  |||||||||||||||  |||||||||  |||
Db       2224 ACGCGCAGCAGGCGGGCCTGCGCCTCCAGCGGCAGCTCGCCGATCTCGTCGAGGAACAGG 2165

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCG 704
              || ||||| || ||||| |||  ||| || || ||  | |||   |||||  |  ||
Db       2164 GTGCCGCCGTCGGCCGCTTCGACCAGGCCGGCGCGGCCGGCGCTGGCGCCGGTGAAGGCA 2105

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764
              ||   |||||  |||||||| |||  ||||||| |  |||||||| ||||||||  | |
Db       2104 CCTTTCTCGTGGCCGAACAGTTCGGACTCGATCAGGGTCTCGGGGATGGCCGCGCAGTTC 2045

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824
              ||    |  | |||   ||     ||   |   ||   ||||| ||| ||||| | |  |
Db       2044 ACCGAGATCAGCGGTGCCTTGGCGCGACGCGAGAGGTTGTGCAGCGCACGCGCGACCAGC 1985

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTG 884
              |||||||| |||||  |||||||   | ||| |   |  |||  | || || |||||||
Db       1984 TCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACGTTGGAATCGGTGGGAGCGACCTTG 1925

Qy        885 CGCACCTGCGCGAACA 900
              ||| | ||   || | |
Db       1924 CGGATCTTGCCGTAAA 1909


RESULT 9
US-09-252-991A-15531
; Sequence 15531, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER:  US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 15531
;   LENGTH: 1386
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15531

  Query Match             30.5%;  Score 274.8;  DB 3;  Length 1386;
  Best Local Similarity   60.9%;  Pred. No. 4.2e-36;
  Matches  447;  Conservative    0;  Mismatches  287;  Indels    0;  Gaps    0;

Qy        167 GGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGC 226
              |||| |  |||||          |  ||| |   || || |||||||        |||||||
Db        211 GGACAACGGCGGGCCGTCGCCGCGCGCGCATTCCAGGCCGAGGTCCTGGGCCTCGATCTG 270

Qy        227 GCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCC 286
              | ||| | ||  || ||||   ||||  | ||   |||  |||| || ||||| |||||
Db        271 CTTTCCCTCTGCCAGGACCAACCCGCGACGTACCCGGTTGGCCAGTTCGCGGACATTGCC 330

Qy        287 CGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTG 346
              |||||| |    |  ||| || |||||  |||| ||| |    |  | ||||     |
Db        331 CGGCCAGGCGTGCATCGCCATCGCCGCCAGCGCGTCCTCGCTGAAGCGGCGCGGGCGGCG 390

Qy        347 CCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGG 406
              ||  || || |  |          ||      |||| |   ||  |  |||| ||
Db        391 GCCGGTCTCCAGGCTGTAGAGCCGGGAGAAATGCTGGGCCAGCAGCGGCAGGTCCTCGCT 450

Qy        407 CCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCT 466
              ||| || |||| |||||      | || || ||| ||    ||||||  | || |||||
Db        451 CCGTTCGCGCAACGGCGCGGTGCCGACCTGCAGGACGTTGAGCCGGTAGTAGAGGTCCTC 510

Qy        467 CCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCG 526
              |||||  ||||  |    | |  |||| | || ||    || |||||||| | || |||
Db        511 GCGGAACGTCCCGGCATCGATCGCCGCCTCCAGGTCGACATGGGTCGCCGCCAGCACCCG 570

Qy        527 CACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAA 586
              ||||||| ||      |  || |  || ||||||  |  |      ||| ||||
Db        571 CACGTCCACCTCGACCGGCCGGCTACCGCCGACCCGCTCGATCTGCTGTTCCTGGAGAAA 630

Qy        587 ACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGT 646
              |||| ||||     |||||| || || |||||| |||||||  |||||||| | ||| ||
Db        631 GCGCAACAGGTTGGCCTGCAACTCCAGCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGT 690

Qy        647 CCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCC 706
              ||||| || |||||||||  | ||||  ||      |  | |||||| | |   |||||
Db        691 GCCGCCGTCGGCCGCCTCGATCCGCCCGACCTTGCGCTGGTGCGCGCCGGTGAAGGCGCC 750

Qy        707 GGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCAC 766
              |||||  |||||||||||||  || |||| |     ||||| |||||| |||| |  |
Db        751 CTTCTCGTGGCCGAACAGCTCGGACTGGATCAGGTGCTCGGGAATCGCCCCGCAGTTGAT 810

Qy        767 GAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTC 826
              |||||||||  | |   ||| |   | |||||||||   | |||||  | |    ||
Db        811 CGCGACGAACGGCTTGTCCGCGCGCCGCGACTGCCGGTGCAGGGTACGCGCGACCAGTTC 870
```

```
Qy        827 CTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCG 886
              |||||| ||||| | ||||||     | ||| | || ||  ||| |||||||||| |||
Db        871 CTTGCCGGTGCCGCTCTCGCCACGGATCAGTACCGGCGAGTCGGTCGGCGCCAGCTTCGC 930

Qy        887 CACCTGCGCGAACA 900
              || | ||  |  ||
Db        931 CAGCAGCCGGCGCA 944


RESULT 10
US-09-252-991A-15308/c
; Sequence 15308, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 15308
;   LENGTH: 1461
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15308

  Query Match             30.5%;  Score 274.8;  DB 3;  Length 1461;
  Best Local Similarity   60.9%;  Pred. No. 4.2e-36;
  Matches  447;  Conservative    0;  Mismatches  287;  Indels    0;  Gaps    0;

Qy        167 GGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGC 226
              |||  |  |||||       |   ||| |   || || |||||||       |||||||
Db       1311 GGACAACGGCGGGGCCGTCGCCGCGCGCGCATTCCAGGCCGAGGTCCTGGGCCTCGATCTG 1252

Qy        227 GCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCC 286
              | ||| | || || ||||    ||||  | ||   |||  |||| || ||||| |||||
Db       1251 CTTTCCCTCTGCCAGGACCAACCCGCGACGTACCCGGTTGGCCAGTTCGCGGACATTGCC 1192

Qy        287 CGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTG 346
              ||||||  |    | ||| || |||||  |||| ||| |     |    | ||||     |
Db       1191 CGGCCAGGCGTGCATCGCCATCGCCGCCAGCGCGTCCTCGCTGAAGCGGCGCGGGCGGCG 1132

Qy        347 CCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGG 406
              ||  || || |  |      ||      || ||| |   ||  |  |||| ||
Db       1131 GCCGGTCTCCAGGCTGTAGAGCCGGGAGAAATGCTGGGCCAGCAGCGGCAGGTCCTCGCT 1072

Qy        407 CCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCT 466
              ||| || |||| ||||       | || || ||| ||     |||||| | ||| |||||
Db       1071 CCGTTCGCGCAACGGCGCGGTGCCGACCTGCAGGACGTTGAGCCGGTAGTAGAGGTCCTC 1012

Qy        467 CCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCG 526
              |||||  ||||  |  | |  |||| | || ||    || |||||||| | || |||
Db       1011 GCGGAACGTCCCGGCATCGATCGCCGCCTCCAGGTCGACATGGGTCGCCGCCAGCACCCG 952

Qy        527 CACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAA 586
              |||||| ||       |  ||  |   || ||||||   | |        ||| | |||
Db        951 CACGTCCACCTCGACCGGCCGGCTACCGCCGACCCGCTCGATCTGCTGTTCCTGGAGAAA 892

Qy        587 ACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGT 646
              ||||  ||||     |||||| || || |||||| |||||||| | ||| ||
Db        891 GCGCAACAGGTTGGCCTGCAACTCCAGCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGT 832

Qy        647 CCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCCGCCGCGAGCGCGCCCGCGGCCGCGCC 706
              |||||  || |||||||||  | ||||  ||      |  | |||||||  | |   |||||
Db        831 GCCGCCGTCGGCCGCCTCGATCCGCCCGACCTTGCGCTGGTGCGCGCCGGTGAAGGCGCC 772

Qy        707 GGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCAC 766
              |||||  |||||||||||  || |||| |     |||||| |||||| |||| |  |
Db        771 CTTCTCGTGGCCGAACAGCTCGGACTGGATCAGGTGCTCGGGAATCGCCCCGCAGTTGAT 712

Qy        767 GAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTC 826
              ||||||||||  | |   ||| |   | |||||||||  |  |||| | |    ||
Db        711 CGCGACGAACGGCTTGTCCGCGCGCCGCGACTGCCGGTGCAGGGTACGCGCGACCAGTTC 652

Qy        827 CTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCG 886
              |||||| ||||| | ||||||     | ||| | || ||  ||| |||||||||| |||
Db        651 CTTGCCGGTGCCGCTCTCGCCACGGATCAGTACCGGCGAGTCGGTCGGCGCCAGCTTCGC 592

Qy        887 CACCTGCGCGAACA 900
              || | ||  |  ||
Db        591 CAGCAGCCGGCGCA 578


RESULT 11
US-09-902-540-9015/c
; Sequence 9015, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9015
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-9015

  Query Match             29.8%;  Score 267.8;  DB 3;  Length 972;
  Best Local Similarity   60.0%;  Pred. No. 5.6e-35;
  Matches  465;  Conservative    0;  Mismatches  307;  Indels    3;  Gaps    1;

Qy        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
              | || || |   | |  |  | || | ||| |     |   ||||  | |   |||||
Db        814 CATCCAGGGGCGCAGGCGCAGTCGGCGCGTTGACACCCTGAGCGGCTGCGGGAGGCGGCA 755

Qy        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240
              | |   | |     | ||| || | |  ||  |   ||||        ||  |   ||
Db        754 GGGCGGCCCGGTCCACATCGGGCAGGGGCAGGTGCTCTTCGAGAATCTCCCCTTCACAGA 695

Qy        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300
              ||||||||||||  |  |  ||||||||||||||| ||||| || |||||  |
Db        694 GCACCACGGCGCTCTCGATACAGTTCTCCAGCTCCCGCACGTTTCCGGGCCAGCGGTAGC 635

Qy        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360
              | || |  | |   |||  | ||    ||   | |||||| | |||     |  ||| |
Db        634 GCTTGAGGCGCTCCACCGCGGCGGCGCTGAGGCGGGGCGGCGTCAGCCGGTGCCTCCGGG 575

Qy        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGC---ACCGCGTCCCCGGGCCGCTCCCGCA 417
                  |  |           |||||  |||    |   | ||| ||  ||||||||||
Db        574 CGACGGCGGCGACGAAGTGGCGGGCGAGCCGCTCGATGTCCTCCGCGCCGCGCTCCCGCA 515

Qy        418 GCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCC 477
              ||||||||||    | ||| | ||||    | ||| | || |||||     |||||||  |
Db        514 GCGGCGGCAGCACCACCTCGACCACCTTGATGCGGTAGTAGAGGTCCTCGCGGAAGCGGC 455

Qy        478 CCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCC 537
              |||||  ||||||      | || ||||| || || |||||||    |  |||||||| ||
Db        454 CCTCGGCCACCATGCGGGCCAGGTCCCGATGGGTGGCCGCGACGATGCGCACGTCCACCT 395

Qy        538 GCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGC 597
              | |||  | ||     ||||| |||    |  ||||     ||| |  |  |||||||
Db        394 TCACGGCCTGGGTGCCTCCCACGCGCTCGAACTCGCGATCCTGGATGACCCGCAGCAACT 335
```

```
Qy        598 GCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCG 657
              |||||||||  ||  |||||||||||| | ||||||||  | ||  ||| |||||    |
Db        334 TGCCCTGCACCGGCAGGGGCAGCTCGCCAATCTCGTCGATGAACACCGTGCCGCCGCTGG 275

Qy        658 CCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACC 717
              ||||||||  ||   ||  | |        |  ||||||  | |   ||||||   ||||  |
Db        274 CCGCCTCGAACTTGCCGGGCACGCGGTGGTCCGCGCCGGTGAAGGCGCCGCGTTCGTGGC 215

Qy        718 CGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACG 777
              ||||  ||||||  ||||||  |||   ||||  | |||||||||  | |||    | |||  |
Db        214 CGAAGAGCTCGTTCTCGATGAGCGTGGCGGGCAGCGCCGCGCAGTCCACCTTGATGAAGG 155

Qy        778 GCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGC 837
              ||||  |  ||  ||  ||  |     |  |||  ||  ||  ||  ||  |  |  |||||||||   |||
Db        154 GCTGGTCCCTGCGGGGACCATTCACGTGGACGGCACGGGCGAACAGCTCCTTGCCGCTGC 95

Qy        838 CCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTG 892
              |  |  ||||||    ||  |||||  ||||||  |||  |  |||||  ||||||||||| |
Db         94 CACTCTCGCCGCGCAGCAGCACCGTCGCATCGGTGGGCGCGGCCTTGCGCACCAG 40


RESULT 12
US-09-902-540-5664/c
; Sequence 5664, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5664
;   LENGTH: 1053
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-5664

  Query Match             29.8%;  Score 267.8;  DB 3;  Length 1053;
  Best Local Similarity   61.3%;  Pred. No. 5.6e-35;
  Matches  431;  Conservative    0;  Mismatches  272;  Indels    0;  Gaps    0;

Qy        198 TCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGA 257
              || ||||  |||||       ||| | | |    ||  | |||||  |   |||| || |
Db        848 TCCAGCCACAGGTGCCGAGGCTCCACCACCGGCCCATCCGCGAGGATGGCGGCCCGCTCG 789

Qy        258 ACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGC 317
              |  || ||   |||||| |||||||||||||||| |    |||| | |    |||
Db        788 AGCGCATTGCGCAGCTCGCGGACGTTGCCCGGCCAGGTGAACGCCTCCAGCCGCGCCACG 729

Qy        318 GCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGC 377
              |||||||  ||||     |      |  | |||| | |           ||   |||||||
Db        728 GCCTCCGGTGACACGCGCAAGCCTGGACGACCCAGCTCGTCTCCGATGCGGCGCAGCAGC 669

Qy        378 GACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCC 437
              |||  |  |   ||  |  |||| || |||||| |||||  | ||||||||||
Db        668 AGCTCCGACAACGGCCGCAGGTCCTCGCGCCGCTCGCGCAGGGACGGCAGCCGGATGGGG 609

Qy        438 AGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCG 497
              | |||   |||||||  || |||||   |||||    ||     || |||||||  |
Db        608 AACACGGCCAGCCGGTGGTAGAGGTCCTCGCGGAACTCACCGCGCGCCATCATCGCCTTG 549

Qy        498 AGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCC 557
              |||||  ||||  ||  |||||  |  |   ||||||| ||    | |||  |||||    |||
Db        548 AGATCGCGGTTGGTGGCCGCCACCCAGCGCACGTCGGCTTCCAGGGTCCGCGTACCGCCC 489

Qy        558 ACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGC 617
              || ||    |    ||    || | |  ||||||||     ||||  | ||    |  |
Db        488 ACGCGTTCGAAGCGGCGTTCTTGGAGGACGCGCAGCAGCTTGGCCTGGAGCTCGGCCTTC 429

Qy        618 AGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCC 677
              ||||||||||||||||||  |   |  || || |||| ||||  ||||    | ||   |
Db        428 AGCTCGCCGACCTCGTCCAAGAAGAAGGTGCCTCCCTGCGCCAGCTCGATGCGGCCGCGC 369

Qy        678 CGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATC 737
              |||    || | ||||||  | |  ||||||    |||||  ||||| ||||||| ||| |
Db        368 CGCTGGGCCACCGCGCCGGTGAACGCGCCCTTCTCGTGTCCGAAGAGCTCGCTCTCCAGG 309

Qy        738 TGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTC 797
              |||  || |  | |||||||| |  |||  ||||||  |||   | |   |||  |||
Db        308 AGCGTCTCCGTCAGCGCCGCGCAGTTAACGGCCACGAAAGGCCCGTCCGCGCGCTCACTC 249

Qy        798 AGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGC 857
              | ||||||  || |||||   |  |||||||||  || || | |||  |||||||  ||||
Db        248 CACTGGTGCAGGGCCCGCGCGGTCACCTCCTTGCCGGTCCCGCTCTCCCCCACCAGCAGC 189

Qy        858 AGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              | |||  ||||   | ||||  ||||||  ||||  |    | | ||
Db        188 ACCGTGGCCTGCGTGGGCGACACCTTCCGCAGCGCTTCCACCA 146


RESULT 13
US-09-902-540-1041
; Sequence 1041, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1041
;   LENGTH: 11812
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(11812)
;   OTHER INFORMATION: unsure at all n locations
US-09-902-540-1041

  Query Match             29.8%;  Score 267.8;  DB 3;  Length 11812;
  Best Local Similarity   61.3%;  Pred. No. 5.2e-35;
  Matches  431;  Conservative    0;  Mismatches  272;  Indels    0;  Gaps    0;

Qy        198 TCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGA 257
              || ||||  |||||       ||| | | |    ||  | |||||  |   |||| || |
Db      10964 TCCAGCCACAGGTGCCGAGGCTCCACCACCGGCCCATCCGCGAGGATGGCGGCCCGCTCG 11023

Qy        258 ACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGC 317
              |  || ||   |||||| |||||||||||||||| |    |||| | |    |||
Db      11024 AGCGCATTGCGCAGCTCGCGGACGTTGCCCGGCCAGGTGAACGCCTCCAGCCGCGCCACG 11083

Qy        318 GCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGC 377
              |||||||  ||||     |      |  | |||| | |           ||   |||||||
Db      11084 GCCTCCGGTGACACGCGCAAGCCTGGACGACCCAGCTCGTCTCCGATGCGGCGCAGCAGC 11143

Qy        378 GACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCC 437
              |||  |  |   ||  |  |||| || |||||| |||||  | ||||||||||
Db      11144 AGCTCCGACAACGGCCGCAGGTCCTCGCGCCGCTCGCGCAGGGACGGCAGCCGGATGGGG 11203

Qy        438 AGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCG 497
              | |||   |||||||  || |||||   |||||    ||     || |||||||  |
Db      11204 AACACGGCCAGCCGGTGGTAGAGGTCCTCGCGGAACTCACCGCGCGCCATCATCGCCTTG 11263
```

```
Qy        498 AGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCC 557
              ||||| ||||  || ||||| | |    |||||||| ||    | |||  |||||     |||
Db      11264 AGATCGCGGTTGGTGGCCGCCACCCAGCGCACGTCGGCTTCCAGGGTCCGCGTACCGCCC 11323

Qy        558 ACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGC 617
              || ||    |    ||    || | |  ||||||||     |||| | ||    |  |
Db      11324 ACGCGTTCGAAGCGGCGTTCTTGGAGGACGCGCAGCAGCTTGGCCTGGAGCTCGGCCTTC 11383

Qy        618 AGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCC 677
              ||||||||||||||||  |   |   || || |||| ||||  ||||    | ||   |
Db      11384 AGCTCGCCGACCTCGTCCAAGAAGAAGGTGCCTCCCTGCGCCAGCTCGATGCGGCCGCGC 11443

Qy        678 CGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATC 737
              |||   || | |||||| | |  ||||||   |||||  |||||  |||||||  ||| |
Db      11444 CGCTGGGCCACCGCGCCGGTGAACGCGCCCTTCTCGTGTCCGAAGAGCTCGCTCTCCAGG 11503

Qy        738 TGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTC 797
              |||  || |  | |||||||| |  |||  |||||| |||   | |   |||   |||
Db      11504 AGCGTCTCCGTCAGCGCCGCGCAGTTAACGGCCACGAAAGGCCCGTCCGCGCGCTCACTC 11563

Qy        798 AGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGC 857
              | ||||||  || |||||   | |||||||| || || | ||| |||||| ||||
Db      11564 CACTGGTGCAGGGCCCGCGCGGTCACCTCCTTGCCGGTCCCGCTCTCCCCCACCAGCAGC 11623

Qy        858 AGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              | ||| ||||   | |||| ||||||  |||| |     | | ||
Db      11624 ACCGTGGCCTGCGTGGGCGACACCTTCCGCAGCGCTTCCACCA 11666


RESULT 14
US-09-902-540-963
; Sequence 963, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 963
;   LENGTH: 12849
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-963

  Query Match             29.8%;  Score 267.8;  DB 3;  Length 12849;
  Best Local Similarity   60.0%;  Pred. No. 5.2e-35;
  Matches  465;  Conservative    0;  Mismatches  307;  Indels    3;  Gaps    1;

Qy        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
              | || || |   | |  |  | || | ||| |    |   ||||  | |   |||||
Db      11483 CATCCAGGGGCGCAGGCGCAGTCGGCGCGTTGACACCCTGAGCGGCTGCGGGAGGCGGCA 11542

Qy        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240
              | |    | |      | ||| ||  | |  ||  |  ||||        ||  |   ||
Db      11543 GGGCGGCCCGGTCCACATCGGGCAGGGGCAGGTGCTCTTCGAGAATCTCCCCTTCACAGA 11602

Qy        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300
              |||||||||||||  |   |  ||||||||||||||||| ||||| || ||||||  |
Db      11603 GCACCACGGCGCTCTCGATACAGTTCTCCAGCTCCCGCACGTTTCCGGGCCAGCGGTAGC 11662

Qy        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360
              |  || |  | |   ||| | ||    ||   | |||||| | |||    |  ||| |
Db      11663 GCTTGAGGCGCTCCACCGCGGCGGCGCTGAGGCGGGGCGGCGTCAGCCGGTGCCTCCGGG 11722

Qy        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGC---ACCGCGTCCCCGGGCCGCTCCCGCA 417
                    | |       |||||  |||    |  | |||  ||   ||||||||||
Db      11723 CGACGGCGGCGACGAAGTGGCGGGCGAGCCGCTCGATGTCCTCCGCGCCGCGCTCCCGCA 11782

Qy        418 GCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCC 477
              ||||||||||   | ||| | ||||   | ||| | || |||||    |||||||  |
Db      11783 GCGGCGGCAGCACCACCTCGACCACCTTGATGCGGTAGTAGAGGTCCTCGCGGAAGCGGC 11842

Qy        478 CCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCC 537
              |||||  ||||||     | || ||||| || || |||||||  |  |||||||| ||
Db      11843 CCTCGGCCACCATGCGGGCCAGGTCCCGATGGGTGGCCGCGACGATGCGCACGTCCACCT 11902

Qy        538 GCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGC 597
              |  |||  | ||     ||||| |||   | ||||    |||  |  |||||||
Db      11903 TCACGGCCTGGGTGCCTCCCACGCGCTCGAACTCGCGATCCTGGATGACCCGCAGCAACT 11962

Qy        598 GCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCG 657
              |||||||||  ||  |||||||||| | ||||||||  | || ||| |||||     |
Db      11963 TGCCCTGCACCGGCAGGGGCAGCTCGCCAATCTCGTCGATGAACACCGTGCCGCCGCTGG 12022

Qy        658 CCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACC 717
              |||||||| ||   || | |          | ||||||  | |   ||||||   ||||  |
Db      12023 CCGCCTCGAACTTGCCGGGCACGCGGTGGTCCGCGCCGGTGAAGGCGCCGCGTTCGTGGC 12082

Qy        718 CGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACG 777
              |||| ||||||  ||||||  |||   |||| | |||||||| | |||    | |||  |
Db      12083 CGAAGAGCTCGTTCTCGATGAGCGTGGCGGGCAGCGCCGCGCAGTCCACCTTGATGAAGG 12142

Qy        778 GCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGC 837
              |||| | ||  || || |    ||| || || || || | |  |||||||| |||
Db      12143 GCTGGTCCCTGCGGGGACCATTCACGTGGACGGCACGGGCGAACAGCTCCTTGCCGCTGC 12202

Qy        838 CCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTG 892
              | | ||||||   || ||||| |||||| ||| | |||||  ||||||||||  |
Db      12203 CACTCTCGCCGCGCAGCAGCACCGTCGCATCGGTGGGCGCGGCCTTGCGCACCAG 12257


RESULT 15
US-09-252-991A-9205/c
; Sequence 9205, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 9205
;   LENGTH: 2379
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9205

  Query Match             29.6%;  Score 266;  DB 3;  Length 2379;
  Best Local Similarity   60.5%;  Pred. No. 1.1e-34;
  Matches  437;  Conservative    0;  Mismatches  285;  Indels    0;  Gaps    0;

Qy        179 AGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGC 238
              | ||  | |  || ||| |   ||     ||||   ||      ||||   | | | | |
Db       2085 ACGCTCCAGGCGCAGGTTGGACAGGTGCAGGTCATCCGCGCGGATCAGCCCGTCGCGGCA 2026

Qy        239 GAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGC 298
              || || || ||   |||| | | ||||||||||||| |||| ||| |||||||| |
Db       2025 GATCAGCAAGGTATGGTGGATGACGTTCTCCAGCTCGCGGATGTTCCCCGGCCAGGAATA 1966

Qy        299 CGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCG 358
              ||    |     |  |||||||     || | |  | |  | |  ||||     ||
Db       1965 GGCACGCAACTTCTGCGCCGCCTCCGGGCTCAACTGGCTCTGGCCGTAGCCCAGGCGACG 1906
```

```
Qy        359 CGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAG 418
Db       1905 GCTGTATTCGGCGATGAAGTGGCGGGTCAGCGGCAGGATGTCGCCGGGTCGCTCGCGCAA 1846

Qy        419 CGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCC 478
Db       1845 CGGCGACAGTTCGAGACTGACGACATTGAGCCGGTAGTACAGGTCCTCGCGGAAGTGCCC 1786

Qy        479 CTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCG 538
Db       1785 GGCATTGATCGCCCGCTCCAGTTGCACGTTGGTCGCCGCCAGCACGCGCACGTCGATGGG 1726

Qy        539 CCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCG 598
Db       1725 GATGCTCTTGCGCGACCCCAGGCGCACCACCTCGCGCTCCTGCAACACGCGCAGCAACTT 1666

Qy        599 CCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGC 658
Db       1665 GACCTGGATCGGCATCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGTGCCGCCGTTGGC 1606

Qy        659 CGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCC 718
Db       1605 CTCCTCGAACCAGCCGGCCTTGGCCGCCAGGGCGCCGGTGAAGGCACCCTTCTCGTGGCC 1546

Qy        719 GAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGG 778
Db       1545 GAACAGCTCGGCTTCCACCAGTGACTCGGAGAAGGCCCCGCAGTTGACCGCCACGAACGG 1486

Qy        779 CTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCC 838
Db       1485 GCCGTTGCGCCGCGCGCTGAGGTTGTGGATATGCCGCGCCACCAGTTCCTTGCCGGTGCC 1426

Qy        839 CCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAA 898
Db       1425 GGTCTCGCCGATCACCAGCACGCTGGCCTCGCTCGGCGCCACCTGGCGCAGGTGTTCGAG 1366

Qy        899 CA 900
Db       1365 CA 1364
```

Search completed: April 27, 2006, 12:42:30
Job time : 186.684 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 902.315 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_1_900
Perfect score: 900
Sequence: 1 tgggtatcccgagccgctgg.....cttgcgcacctgcgcgaaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	900	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
2	264.6	29.4	19053	8	US-10-848-111-20	Sequence 20, Appl
c 3	264.4	29.4	1131	7	US-10-282-122A-30192	Sequence 30192, A
c 4	255	28.3	1386	10	US-11-109-593-12	Sequence 12, Appl
c 5	255	28.3	67323	10	US-11-109-593-1	Sequence 1, Appli
c 6	243	27.0	1086	7	US-10-282-122A-30193	Sequence 30193, A
c 7	233	25.9	1104	7	US-10-282-122A-33573	Sequence 33573, A
c 8	231.8	25.8	1512	7	US-10-389-647-75	Sequence 75, Appl
c 9	229.8	25.5	1685	7	US-10-282-122A-23767	Sequence 23767, A
c 10	221.8	24.6	1302	3	US-09-975-719-348	Sequence 348, App
c 11	221.8	24.6	1344	3	US-09-975-719-350	Sequence 350, App
c 12	221.8	24.6	42235	3	US-09-975-719-1	Sequence 1, Appli
c 13	221.2	24.6	2373	8	US-10-411-910A-273	Sequence 273, App
c 14	219.4	24.4	1359	7	US-10-282-122A-11630	Sequence 11630, A
15	218.6	24.3	1749	3	US-09-870-162A-22	Sequence 22, Appl
16	218.6	24.3	14272	3	US-09-870-162A-23	Sequence 23, Appl
c 17	216.8	24.1	1353	7	US-10-282-122A-13776	Sequence 13776, A
c 18	215	23.9	1821	7	US-10-282-122A-20202	Sequence 20202, A
c 19	213.4	23.7	1101	7	US-10-282-122A-31705	Sequence 31705, A
c 20	210.4	23.4	2079	7	US-10-282-122A-39114	Sequence 39114, A
c 21	207.8	23.1	1350	7	US-10-282-122A-13544	Sequence 13544, A
22	206.4	22.9	1836	3	US-09-997-664-98	Sequence 98, Appl
23	206.4	22.9	1836	6	US-10-464-952-98	Sequence 98, Appl
24	206.4	22.9	6491	3	US-09-997-664-1	Sequence 1, Appli
25	206.4	22.9	6491	6	US-10-464-952-1	Sequence 1, Appli
c 26	202.4	22.5	2079	7	US-10-282-122A-39771	Sequence 39771, A
27	195.6	21.7	12591	5	US-10-213-878-6	Sequence 6, Appli
28	195.6	21.7	12591	5	US-10-214-059-6	Sequence 6, Appli
c 29	193.2	21.5	1278	7	US-10-282-122A-7351	Sequence 7351, Ap
c 30	187.6	20.8	2079	7	US-10-282-122A-7099	Sequence 7099, Ap
c 31	187.4	20.8	2373	8	US-10-411-910A-166	Sequence 166, App
c 32	186	20.7	1326	7	US-10-282-122A-31403	Sequence 31403, A
33	176.2	19.6	10404	9	US-10-893-671-91	Sequence 91, Appl
34	176.2	19.6	19528	8	US-10-612-594-9	Sequence 9, Appli
c 35	174.6	19.4	1454	9	US-10-450-763-21322	Sequence 21322, A
36	172	19.1	536165	3	US-09-939-964-1	Sequence 1, Appli
c 37	171.4	19.0	1459	9	US-10-450-763-9193	Sequence 9193, Ap
c 38	167.4	18.6	3363	3	US-09-795-483-1	Sequence 1, Appli
c 39	167.4	18.6	3363	9	US-10-997-917-1	Sequence 1, Appli
c 40	161.4	17.9	1326	7	US-10-282-122A-33737	Sequence 33737, A
c 41	158.2	17.6	6960	5	US-10-194-163-354	Sequence 354, App
c 42	156.8	17.4	1474	7	US-10-282-122A-36621	Sequence 36621, A
43	155.8	17.3	522	3	US-09-975-719-354	Sequence 354, App
c 44	154.4	17.2	3025	5	US-10-194-163-694	Sequence 694, App
c 45	154.2	17.1	636	9	US-10-450-763-7173	Sequence 7173, Ap

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 900; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 2.2e-187;
Matches 900; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TGGGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGGGTATCCCGAGCCGCTGGCGCCGTTCCCACAAGGCCTTGCGGCTGATGCCGAGCCGAC 60

Qy         61 GGGCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGCAATCTCGGTCTCCGTCAGCTCGTCCTGGTGCTCCAGCACGAAGCGGCGGAAATAGC 120

Qy        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
```

```
Qy        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240

Qy        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300

Qy        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360

Qy        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCG 420

Qy        421 GCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCT 480

Qy        481 CGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCC 540

Qy        541 GGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCC 600

Qy        601 CCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCG 660

Qy        661 CCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGA 720

Qy        721 ACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCT 780

Qy        781 GCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCC 840

Qy        841 CCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
```

RESULT 2
US-10-848-111-20
; Sequence 20, Application US/10848111
; Publication No. US20040235107A1
; GENERAL INFORMATION:
; APPLICANT: Rosenberg, Eugene
; APPLICANT: Ron, Eliora
; APPLICANT: Orr, Elisha
; APPLICANT: Paitan, Yossi
; TITLE OF INVENTION: GENE CLUSTER
; FILE REFERENCE: 27757
; CURRENT APPLICATION NUMBER: US/10/848,111
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 19053
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-10-848-111-20

Query Match 29.4%; Score 264.6; DB 8; Length 19053;
Best Local Similarity 59.7%; Pred. No. 7.3e-49;

Matches 463; Conservative 0; Mismatches 309; Indels 3; Gaps 1;

```
Qy        121 CCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCGGAGGAGACGGGCGGAG 180
              | ||| || |  | | |  | || | ||| |    |  ||||  | |  |||||
Db       3271 CATCCAGGGGCGCAGGCGCAGTCGGCGCGTTGACACCCTGAGCGGCTGCGGGAGGCGGCA 3330

Qy        181 GCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGA 240
              || | | |  | ||| || | |  || |  |||  |  || |  ||
Db       3331 GGGCGGCCCGGTCCACATCGGGCAGGGGCAGGTGCTCTTCGAGAATCTCCCCTTCACAGA 3390

Qy        241 GCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCG 300
              |||||||||||  |  |  ||||||||||||||| ||||| || |||||  |
Db       3391 GCACCACGGCGCTCTCGATACAGTTCTCCAGCTCCCGCACGTTTCCGGGCCAGCGGTAGC 3450

Qy        301 CCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCG 360
              |  || |  | |  ||| | ||   ||   | |||||| | |||   |  ||| |
Db       3451 GCTTGAGGCGCTCCACCGCGGCGGCGCTGAGGCGGGGCGGCGTCAGCCGGTGCCTCCGGG 3510

Qy        361 TCCGTCGCTTCAGCAGCGACTCGGCGATGCGC---ACCGCGTCCCCGGGCCGCTCCCGCA 417
              |  |  ||||| |||  |  | ||| ||  ||||||||||
Db       3511 CGACGGCGGCGACGAAGTGGCGGGCGAGCCGCTCGATGTCCTCCGCGCCGCGCTCCCGCA 3570

Qy        418 GCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCC 477
              |||||||||||  | ||| | ||||  |  ||| | || |||||  |||||||  |
Db       3571 GCGGCGGCAGCACCACCTCGACCACCTTGATGCGGTAGTAGAGGTCCTCGCGGAAGCGGC 3630

Qy        478 CCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCC 537
              |||||  ||||||     | || ||||| || || |||||||  |  |||||||||  ||
Db       3631 CCTCGGCCACCATGCGGGCCAGGTCCCGATGGGTGGCCGCGACGATGCGCACGTCCACCT 3690

Qy        538 GCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGC 597
              |  |||  | ||   ||||| |||   |  ||||   ||| |  |  |||||||
Db       3691 TCACGGCCTGGGTGCCTCCCACGCGCTCGAACTCGCGATCCTGGATGACCCGCAGCAACT 3750

Qy        598 GCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCG 657
              |||||||||  ||  |||||||||||  | ||||||||  | ||  ||  |||||   |
Db       3751 TGCCCTGCACCGGCAGGGGCAGCTCGCCAATCTCGTCGATGAACACGGTGCCGCCGCTGG 3810

Qy        658 CCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACC 717
              |||| ||| ||   ||  | |          | ||||||  | |   ||||||   ||||  |
Db       3811 CCGCTTCGAACTTGCCGGGCACGCGGTGGTCCGCGCCGGTGAAGGCGCCGCGTTCGTGGC 3870

Qy        718 CGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACG 777
              |||| ||||||  ||||||  |||   |||| | |||||||| | |||   | ||| |
Db       3871 CGAAGAGCTCGTTCTCGATGAGCGTGGCGGGCAGCGCCGCGCAGTCCACCTTGATGAAGG 3930

Qy        778 GCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGC 837
              |||| | ||  ||  || |  |  ||| || || || || | |  |||||||  |||
Db       3931 GCTGGTCCCTGCGGGGACCATTCACGTGGACGGCACGGGCGAACAGCTCCTTGCCGCTGC 3990

Qy        838 CCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTG 892
              | | ||||||   || ||||| |||||| ||| | |||||  ||||||||||| |
Db       3991 CACTCTCGCCGCGCAGCAGCACCGTCGCATCGGTGGGCGCGGCCTTGCGCACCAG 4045
```

RESULT 3
US-10-282-122A-30192/c
; Sequence 30192, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30192
;   LENGTH: 1131
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30192

  Query Match             29.4%;  Score 264.4;  DB 7;  Length 1131;
  Best Local Similarity   60.4%;  Pred. No. 1.4e-48;
  Matches  436;  Conservative    0;  Mismatches  286;  Indels    0;  Gaps    0;

Qy        179 AGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGC 238
              | ||  | |  || | | |   ||     ||||   ||      ||||   | | | |
Db        837 ACGCTCCAGGCGCAGATTGGACAGGTGCAGGTCATCCGCGCGGATCAGCCCGTCGCGGCA 778

Qy        239 GAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGC 298
              || || || ||   |||| | | ||||||||||||||| |||| ||| ||||||||
Db        777 GATCAGCAAGGTATGGTGGATGACGTTCTCCAGCTCGCGGATGTTCCCCGGCCAGGAATA 718

Qy        299 CGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCG 358
              ||     |   |  ||||||||   || |  | | | |  ||||    ||
Db        717 GGCACGCAACTTCTGCGCCGCCTCCGGGCTCAACTGGCTCTGGCCGTAGCCCAGGCGACG 658

Qy        359 CGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAG 418
              |   |      |  |      ||  |   |||    ||| ||||| |||||  ||||
Db        657 GCTGTATTCGGCGATGAAGTGGCGGGTCAGCGGCAGGATGTCGCCGGGTCGCTCGCGCAA 598

Qy        419 CGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCC 478
              |||||  |||   ||      | ||    |||||| | ||||||||     ||||||  |||
Db        597 CGGCGACAGTTCGAGACTGACGACATTGAGCCGGTAGTACAGGTCCTCGCGGAAGTGCCC 538

Qy        479 CTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCG 538
              |    | |  |  | | || | |  ||  |||||||| | ||  |||||||     |
Db        537 GGCATTGATCGCCCGCTCCAGTTGCACGTTGGTCGCCGCCAGCACGCGCACGTCGATGGG 478

Qy        539 CCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCG 598
                |       |||||||  |||  ||| |||| |  |||||| |  |||||||
Db        477 GATGCTCTTGCGCGACCCCAAGCGCACCACCTCGCGCTCCTGCAACACGCGCAGCAACTT 418

Qy        599 CCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGC 658
              ||||  | |  ||||||||| ||||||| ||||||||  | ||| || |||||  |  ||
Db        417 GACCTGGATCGGCATCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGTGCCGCCGTTGGC 358

Qy        659 CGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCC 718
              | ||||| ||   || |||   |||||  || |||||  | |   |||||  |||||  ||
Db        357 CTCCTCGAACCAGCCGGCCTTGGCCGCCAGGGCGCCGGTGAAGGCGCCCTTCTCGTGGCC 298

Qy        719 GAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGG 778
              |||||||||| | || | | | |  |||| ||  ||| |||| |  ||   |||||||||
Db        297 GAACAGCTCGGCTTCCACCAGTGACTCGGAGAAGGCCCCGCAGTTGACCGCCACGAACGG 238

Qy        779 CTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCC 838
                    ||||||| ||| ||  ||| |     ||||||| |   |||||||| |||||
Db        237 ACCGGTGCGCCGCGCGCTGAGGTTGTGGATATGCCGCGCCACCAGTTCCTTGCCGGTGCC 178

Qy        839 CCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAA 898
                ||||||  | |||||||||    | ||||||||||||||||||| |||||   ||  |||
Db        177 GGTCTCGCCGATCACCAGCACGCTGGCCTCGCTCGGCGCCACCTGGCGCAGGTGTTCGAG 118

Qy        899 CA 900
              ||
Db        117 CA 116


RESULT 4
US-11-109-593-12/c
; Sequence 12, Application US/11109593
; Publication No. US20050233369A1
; GENERAL INFORMATION:
;  APPLICANT: Reid, Ralph
;  APPLICANT:  Reeves, Christopher D.
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR JERANGOLIDS
;  FILE REFERENCE: 010108.01
;  CURRENT APPLICATION NUMBER: US/11/109,593
;  CURRENT FILING DATE:  2005-04-18
;  PRIOR APPLICATION NUMBER: US 60/563,843
;  PRIOR FILING DATE: 2004-04-19
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 12
;   LENGTH: 1386
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-109-593-12

  Query Match             28.3%;  Score 255;  DB 10;  Length 1386;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-46;
  Matches  420;  Conservative    0;  Mismatches  275;  Indels    0;  Gaps    0;

Qy        206 GAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTT 265
              ||| | | |   |  ||||     |  ||  |||||| | ||||| || |  | |  |||
Db       1152 GAGATCGTCGACCGTGATCTGATCGTACCGCGCGAGCGCGACGGCCCGCTCGATGCAGTT 1093

Qy        266 CTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGC 325
              ||  ||||| |||||||||||| |||||||      ||  |||     | ||  |||   |||
Db       1092 CTGGAGCTCGCGGACGTTGCCGGGCCACGCGTAGGCGACGAGCTTCTCGGCCGCGGGCGC 1033

Qy        326 CGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGC 385
              ||| ||  |        | | || || | ||   ||        |||  || |    ||  ||
Db       1032 CGAGAGGCCCTTGATGGGCCGCTCCTTGACGGTCGCGAAGTGCTCGAGGAAGCGCTGCGC 973

Qy        386 GATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCG 445
              ||   ||| | ||||  ||   ||   || ||||||||||||   ||  |   | |||
Db        972 GAGCAGCAGCACGTCGGCGCCGCGGGAGCGGAGCGGCGGCAGATCGACGTGGACCACGTT 913

Qy        446 CAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCG 505
              | |||| |||| |||||     |||||||  | |||     | |   | | | || || ||
Db        912 GATCCGGTAATAGAGGTCCTCGCGGAAGCGGCGCTCCTCGATCGCGGTCTCCAGGTCGCG 853

Qy        506 GTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCG 565
              ||  |||||||||| ||||||||||||  |   |       ||  || ||  ||||
Db        852 GTTGGTCGCCGCGATGAGCCGCACGTCCACGGGGACCTCCTCGTCCGCGCCGACGGGCCG 793

Qy        566 CACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCC 625
              ||  |||  |  |||  |     ||||| ||    |  |||  | |   | ||||||||||||
Db        792 GACGACGCGCTCCTGGAGGGCGCGCAGGAGCTTCGGCTGGAGCCCGACCGGCAGCTCGCC 733

Qy        626 GACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGC 685
              || |||||||| | ||| || |||||  |  |||  | || |||| ||   || |||| |
Db        732 GATCTCGTCGAGGAACAGGGTGCCGCCGTGGGCCCGCGCGAACAGGCCGCTCCTCGCCCC 673
```

```
Qy        686 GAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATC 745
              ||| | | |   |||||   | |||  ||||||||||||  ||||| | | || ||
Db        672 CTTGGCGTCGGTGAAGGCGCCCCGCGCGTGGCCGAACAGCTCGGTCTCGAGCAGGGCGTC 613

Qy        746 GGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTG 805
              |||| |||||||||| | |||    | |||||    ||| |  ||| ||||  || ||||
Db        612 CGGGACCGCCGCGCAGTTCACCGCGATGAACGCGCCCTGGCCGCGCCGGCTGCGCTGGTG 553

Qy        806 CACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGC 865
              ||| ||||||||  |    |||||||| ||||| | ||||||    |  || | ||  |
Db        552 CACGGCGCGCGCGACGAGCTCCTTGCCGGTGCCGCTCTCGCCGGTGATGAGGATCGAGGT 493

Qy        866 CTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||      || |||  |   | | |  |||| |
Db        492 GTCCGAGTCGGCGACCCGGTCGAGCAGATCGAAGA 458


RESULT 5
US-11-109-593-1/c
; Sequence 1, Application US/11109593
; Publication No. US20050233369A1
; GENERAL INFORMATION:
;  APPLICANT: Reid, Ralph
;  APPLICANT:  Reeves, Christopher D.
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR JERANGOLIDS
;  FILE REFERENCE: 010108.01
;  CURRENT APPLICATION NUMBER: US/11/109,593
;  CURRENT FILING DATE:  2005-04-18
;  PRIOR APPLICATION NUMBER: US 60/563,843
;  PRIOR FILING DATE: 2004-04-19
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 67323
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-109-593-1

  Query Match             28.3%;  Score 255;  DB 10;  Length 67323;
  Best Local Similarity   60.4%;  Pred. No. 7e-47;
  Matches  420;  Conservative    0;  Mismatches  275;  Indels    0;  Gaps    0;

Qy        206 GAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTT 265
              ||| | | |   |  ||||     |  ||  |||||| | ||||| || |  | |  |||
Db       9791 GAGATCGTCGACCGTGATCTGATCGTACCGCGCGAGCGCGACGGCCCGCTCGATGCAGTT 9732

Qy        266 CTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGC 325
              ||  ||||| |||||||||||| |||||||     ||  |||    | ||  |||   |||
Db       9731 CTGGAGCTCGCGGACGTTGCCGGGCCACGCGTAGGCGACGAGCTTCTCGGCCGCGGGCGC 9672

Qy        326 CGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGC 385
              ||| ||  |       | | || || | ||   ||       |||  || |    ||  ||
Db       9671 CGAGAGGCCCTTGATGGGCCGCTCCTTGACGGTCGCGAAGTGCTCGAGGAAGCGCTGCGC 9612

Qy        386 GATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCG 445
              ||   ||| | ||||  ||   ||    || ||||||||||   ||  |  | |||
Db       9611 GAGCAGCAGCACGTCGGCGCCGCGGGAGCGGAGCGGCGGCAGATCGACGTGGACCACGTT 9552

Qy        446 CAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCG 505
              | |||| |||| |||||     |||||||  | |||    |  |  | | || || |||
Db       9551 GATCCGGTAATAGAGGTCCTCGCGGAAGCGGCGCTCCTCGATCGCGGTCTCCAGGTCGCG 9492

Qy        506 GTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCG 565
              ||  |||||||||| |||||||||||| |  |      |    ||  || ||  ||||
Db       9491 GTTGGTCGCCGCGATGAGCCGCACGTCCACGGGGACCTCCTCGTCCGCGCCGACGGGCCG 9432

Qy        566 CACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCC 625
              ||  ||| |  ||| |     ||||| ||   |  ||| | |   | ||||||||||||
Db       9431 GACGACGCGCTCCTGGAGGGCGCGCAGGAGCTTCGGCTGGAGCCCGACCGGCAGCTCGCC 9372

Qy        626 GACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGC 685
              || |||||||||  | ||| || ||||| |  |||  | || |||| ||    || |||| |
Db       9371 GATCTCGTCGAGGAACAGGGTGCCGCCGTGGGCCCGCGCGAACAGGCCGCTCCTCGCCCC 9312
```

```
Qy        686 GAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATC 745
              ||| | | |   |||||   | |||  ||||||||||||  ||||| | | || ||
Db       9311 CTTGGCGTCGGTGAAGGCGCCCCGCGCGTGGCCGAACAGCTCGGTCTCGAGCAGGGCGTC 9252

Qy        746 GGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTG 805
              |||| |||||||||| | |||    | |||||    ||| |  ||| ||||  || ||||
Db       9251 CGGGACCGCCGCGCAGTTCACCGCGATGAACGCGCCCTGGCCGCGCCGGCTGCGCTGGTG 9192

Qy        806 CACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGC 865
              ||| ||||||||  |    |||||||| ||||| | ||||||    |  || | ||  |
Db       9191 CACGGCGCGCGCGACGAGCTCCTTGCCGGTGCCGCTCTCGCCGGTGATGAGGATCGAGGT 9132

Qy        866 CTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||      || |||  |   | | |  |||| |
Db       9131 GTCCGAGTCGGCGACCCGGTCGAGCAGATCGAAGA 9097


RESULT 6
US-10-282-122A-30193/c
; Sequence 30193, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30193
;   LENGTH: 1086
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30193

  Query Match             27.0%;  Score 243;  DB 7;  Length 1086;
  Best Local Similarity   57.9%;  Pred. No. 6.5e-44;
  Matches  449;  Conservative    0;  Mismatches  325;  Indels    1;  Gaps    1;

Qy        106 AGCGGCGGAAATAGCCCTCGAGCGAGTCCGAAGGCGGCGCCCCGTCGCGCAGCGATGCG- 164
```

```
Db        887 AACAGCGCCTGCAGCGCCGCCTCCAACGCCGAGCCGGCGGCGCTTTCGGGAGCGCGGCGC 828

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224

Db        827 AAGCCGTTCGGCTCCTGGCGCGCACGCATGTCGATGAGGTGCAGGTCGGCCGGCTGGATC 768

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284

Db        767 AGACGGTTGCGGCAGACCAGCAGCGCGTGGTGGATGGCGTTCTCCAGCTCGCGGATATTG 708

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344

Db        707 CCCGCCCAGGAATGCCCGAGCAGCTTGCGCTCGGCTTCCGGGCTCAGGCTCGCGCTGGTG 648

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404

Db        647 TAGCCGAGGCGCCGGCAGTGCTCGACGATGAAGTACTCGGCCAGCGGCAGGATATCGCCG 588

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464

Db        587 CGGCGCTCGCGCAACGGCGGCAGTTGGATGGTCGCCACGTGCAGGCGGTAGAACAGGTCC 528

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524

Db        527 TCGCGGAAGTGTCCGGCGACCACCGCGTCAGCCAGGTCGACGTTGGTCGCCGCCACCAGC 468

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAA 584

Db        467 CGCACGTCGATGGGGATCGGCCGGCGCGAACCGAGCCGCACCACCTCGCGTTCCTGCAGC 408

Qy        585 AAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGC 644

Db        407 ACGCGCAGCAGCTTGACCTGGATGCTCGGCGGCAGGTCGCCGATCTCGTCGAGGAACAGG 348

Qy        645 GTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCCGCCGCGAGCGCGCCCGCGGCCGCG 704

Db        347 GTGCCGCCGTTGGCCGCCTCGAACCAGCCGGCCTTGCTCTGCAACGCACCGGTAAAGGCG 288

Qy        705 CCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGC 764

Db        287 CCCTTCTCGTGGCCGAACAGCTCGCTCTCCACCAGGCTCTCGGCGAAAGCACCGCAGTTG 228

Qy        765 ACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTC 824

Db        227 ACCGCGACGAACGGGCCGCCGTTGCGTCCGCTGAGGGCGTGGATATGGCGCGCCACCAGC 168

Qy        825 TCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCA 879

Db        167 TCCTTGCCGGTGCCGGTCTCGCCGATGACCAGCGCGTTGGCCTCGCTGGGCGCCA 113


RESULT 7
US-10-282-122A-33573/c
; Sequence 33573, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33573
;   LENGTH: 1104
;   TYPE: DNA
;   ORGANISM: Pseudomonas syringae
US-10-282-122A-33573

  Query Match             25.9%;  Score 233;  DB 7;  Length 1104;
  Best Local Similarity   57.5%;  Pred. No. 9.7e-42;
  Matches  419;  Conservative    0;  Mismatches  310;  Indels    0;  Gaps    0;

Qy        172 CGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGC 231

Db        826 CGTCCTGACGCTCAATCCTCAAGTTGGACATGCGCAGGTCATCACGCTGAATGATCCCGT 767

Qy        232 CCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCC 291

Db        766 CGCGGCAGACCAGCAACGTGTGGTGAATGACGTTTTCCAGCTCGCGAATGTTGCCCGGCC 707

Qy        292 ACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCA 351

Db        706 AACTGTAGGTCCTGAGTTTGTGCTCGGCCTCGGCACTGATCCGCACGGGGCCATGGCGCA 647

Qy        352 TCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCT 411

Db        646 GGCGCTGGCTGTAGACCTCGATGAAATGCCGGGCCAGTGGCAGGATATCGCCCGGCCGGT 587

Qy        412 CCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGA 471

Db        586 CACGCAACGGGCTCAGTTCCAGATTGACCACGTCCAGCCGGTAATACAGGTCCTCGCGGA 527

Qy        472 AGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGT 531

Db        526 AATGCCCGGCATTGATGGCTTTTTCCAGCTGCACGTTGGTGGCCGCCAGCACCCGCACAT 467

Qy        532 CCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCA 591

Db        466 CAATCGGAATGCTCTTGCGCGAGCCGAGGCGAACCACTTCGCGCTCTTGAAGCACGCGCA 407

Qy        592 GCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGC 651

Db        406 GCAGTTTGACCTGCAGAGTCATGGGCAGATCGCCGATCTCATCGAGAAACAGCGTGCCGC 347

Qy        652 CCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCT 711

Db        346 CGTTGGCCTCTTCAAACCAGCCGGCCTTGGCGCTGATTGCACCGGTAAACGCGCCTTTGT 287

Qy        712 CGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCA 771

Db        286 CGTGGCCGAACAGCTCGGCCTCGATCAGGTTTTCGGAGAACGCGCCGCAGTTGACCGCCA 227
```

```
Qy        772 CGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGC 831
              | ||||||| |  || || | ||||| |  ||| ||     |  || | |  |||||||
Db        226 CAAACGGCTTGTCACGTCGGGCGCTCAACTCGTGAACGTGCCTTGCAACCAGTTCCTTGC 167

Qy        832 CCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCT 891
              | |||||   |||||| |  | |||||   | || ||||| || ||||| | ||||||| |
Db        166 CGGTGCCGGTCTCGCCGATGATCAGCACGCTGGCGTCGCTGGGTGCCACTTGGCGCACAT 107

Qy        892 GCGCGAACA 900
              | |||| ||
Db        106 GAGCGAGCA 98
```

RESULT 8
US-10-389-647-75/c
; Sequence 75, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Candi
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-75

Query Match 25.8%; Score 231.8; DB 7; Length 1512;
Best Local Similarity 57.5%; Pred. No. 1.7e-41;
Matches 416; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

```
Qy        170 GACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCT 229
              || ||||    ||   ||| ||  ||   ||    | |||  |   |   | |
Db       1344 GAGCGGCGCCTCCGCTTCCGGGGGCGCGCGCAGGGAAAAGTGTTCGGCCAGCAGCATCTC 1285

Qy        230 GCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGG 289
              |||  ||  |||||  || ||||  |  ||       || ||||||||| |||||||| ||
Db       1284 GCCTTCGCAGAGCAGTACCGCGCATTCCACCAATCCCTTCAGCTCCCGAACGTTGCCGGG 1225

Qy        290 CCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCC 349
                |        |  |||       |  |||| ||||| |||    | |||||||  | |
Db       1224 AAAGCCATAGCCGTCGAGCAGTTCCAGCGCTTCCGCGGACCAGCCCAGCGGCGCGCGGCG 1165

Qy        350 CATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCG 409
              ||  | |          ||||    |   |  ||| |     |  ||||| ||    ||
Db       1164 CAGGAACACGCAGGCGCGCTCGGCGAAATGCCGGGCCAGTTCGAGGGCGTCGCCCTCGCG 1105

Qy        410 CTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCG 469
              ||  |||| |||||||| |  |||    | |  |  ||||||| | || |||||    ||
Db       1104 CTGGCGCAACGGCGGCAACTCGATGGGAAACTGCGCCAGCCGGTAGTAGAGGTCTTCGCG 1045

Qy        470 GAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCAC 529
              |||  | || ||||| |||| |    |||| || ||||| || || |||||  |  || ||
Db       1044 GAACTTGCCATCGCGTACCAGCAGGGCGAGGTCACGGTGGGTAGCGGCGATGATGCGTAC 985

Qy        530 GTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACG 589
                ||  ||   | |     |   || | | || |  |||||   |||||| |  ||
Db        984 ATCGACCTTGTGCGTATCGTTGCCGCCGAGCGGGCGAATCTCGCCTTCCTGCAACACGCG 925

Qy        590 CAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCC 649
              || |||     |||| | |   |  ||||  |||||||  |||||||||||||| || ||
Db        924 CAACAGCTTGGCCTGGAGCGAGAGGGGCATGTCGCCGATTTCGTCGAGCAGCAGGGTTCC 865
```

```
Qy        650 GCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGC 709
              ||| ||||| || ||| ||| ||| |   ||  |     || ||  |      |||||
Db        864 GCCATCCGCGGCATCGAACAACCCGGGACGATCGCGCTCGGCACCGGTAAAGGCGCCCTT 805

Qy        710 CTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAG 769
                 ||| ||||||||||||||  || | |    || || | |||||||| | |    |
Db        804 GCGGTAGCCGAACAGCTCGCTTTCCAGCAGGTTTTCCGGAAACGCCGCGCAGTTCTGCAC 745

Qy        770 CACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTT 829
              |||||||| |||     ||||||||||      | ||| | ||| || || | |  ||||||
Db        744 CACGAACGCCTGTGCACGCCGCGGGCCGTACTCGTGGATCGCCCGTGCGACCACCTCCTT 685

Qy        830 GCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCAC 889
              ||| |||||   |||||||  || || ||  ||       | |  ||   |||||||  |
Db        684 GCCGGTGCCGGTCTCGCCCTGCAGCAACACGGTATAGGGGTTGTGCAGGACCTTGCTGAT 625

Qy        890 CTG 892
              | |
Db        624 CAG 622
```

RESULT 9
US-10-282-122A-23767/c
; Sequence 23767, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23767
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23767

Query Match 25.5%; Score 229.8; DB 7; Length 1685;
Best Local Similarity 57.2%; Pred. No. 4.5e-41;

```
  Matches  417;  Conservative    0;  Mismatches  312;  Indels    0;  Gaps    0;

Qy        172 CGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGC 231
              | |||||   ||||      |  |   ||  || |  | |||    ||  |   || |||
Db       1512 CCGGCGTCCGCGGGGCTTCCACGATATCCCGCTCCGGCAGGGAGAGCTGCAGTACGTTGC 1453

Qy        232 CCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCC 291
              | | || |||||  || || || |   |||| |||||||| || ||||||||||||||
Db       1452 CGCGGGTGAGCAATACCGCCCGTTCGATAACGTTTTCCAGCTCGCGCACGTTGCCCGGCC 1393

Qy        292 ACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCA 351
              ||  |     ||  |  |     | | |||||||  |    |          ||||||
Db       1392 ACTCCATGCGCGTCAGAGTGCGCAGGGTTTCCGCCGGAATGCTGTCGATATTGCGCCCCA 1333

Qy        352 TCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCT 411
              |    || |   |          | || ||   | |  ||       ||  |||| || |
Db       1332 TGCGGCGGGCGATTTTAAAGGTAAACGCCTTCACCAGCAGCGGGATATCGTCGGGACGTT 1273

Qy        412 CCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGA 471
              | |||||||||||||||  |||||   | |||    ||||||  | || || ||||| ||||
Db       1272 CGCGCAGCGGCGGCAGGTGGATCGGGAACACGTTGAGCCGGTAGTAGAGATCGCTGCGGA 1213

Qy        472 AGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGT 531
              |   ||  |||   ||||||  |  ||||||  ||||   ||  || |||||||   || ||||
Db       1212 ATTCCCTGTCGATAACCATCTGCTTGAGATCGCGGTTGGTGGCGGCGATCAATCGAACGT 1153

Qy        532 CCGCCCGCCGGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCA 591
              | | | |          |  |||||  ||    |  ||   | |||  |  |  ||||
Db       1152 CGGTCTGAATCAGCTTGTTGCTCCCCAGACGTTCGAACTCCTGCTCCTGGAGGACGCGCA 1093

Qy        592 GCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGC 651
              ||||      |||||  | ||  |||||  ||||| || || |||||  | |||  |  |
Db       1092 GCAGTTTGGGCTGCAGTTCCAGCGGCATATCGCCCACTTCATCGAGGAACAGGGAGCTTT 1033

Qy        652 CCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCT 711
              ||||||   ||| |  | |||    |||  |||   || || | |  |||||||    |
Db       1032 TGTCCGCCAGTTCGAAGCGGCCGATACGCTGGGCGCTGGCCCCGGTGAACGCGCCGCGTT 973

Qy        712 CGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCA 771
              |||  |||||||| |||| ||| | | || |  |||| |||||  |||||  | ||    ||
Db        972 CGTGGCCGAACAGATCGCTCTCCAGCAGCCCGGCGGGCATCGCGGCGCAGTTCATCTTCA 913

Qy        772 CGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGC 831
              | | | || |        ||  ||| ||  ||| |  ||||| || |     ||| ||||
Db        912 CCATCCGCCGACCGTTGCGTCCGCTGAGGTTGTGGATAGCGCGGGCGATAAGCTCTTTGC 853

Qy        832 CCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCT 891
              | |||||    |||||  |  | |||||  ||     |||||| ||||||| |    ||||
Db        852 CGGTGCCGGTTTCGCCGAGGATCAGCACGGTGCTGTCGCTCTGCGCCACCATCTCTACCT 793

Qy        892 GCGCGAACA 900
              ||  || ||
Db        792 GCTTGAGCA 784


RESULT 10
US-09-975-719-348/c
; Sequence 348, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
;  APPLICANT: Ausubel, Frederick M.
;  APPLICANT:  Rahme, Laurence G.
;  TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE: 00786/361003
;  CURRENT APPLICATION NUMBER: US/09/975,719
;  CURRENT FILING DATE:  2001-10-10
;  PRIOR APPLICATION NUMBER: US 09/199,637
;  PRIOR FILING DATE: 1998-11-25
;  PRIOR APPLICATION NUMBER: US 60/066,517
;  PRIOR FILING DATE: 1997-11-25
;  NUMBER OF SEQ ID NOS: 437
;  SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 348
;   LENGTH: 1302
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-975-719-348

  Query Match             24.6%;  Score 221.8;  DB 3;  Length 1302;
  Best Local Similarity   58.3%;  Pred. No. 2.6e-39;
  Matches  425;  Conservative    0;  Mismatches  302;  Indels    2;  Gaps    2;

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              | || | |   | |    | | |||||  ||| | || |  |||| |      || ||||
Db       1133 GCGGCGCCTTCCTGGCTGGCACCCGGCGCATCGGCCAGGCGCAGGTCGTGAGGCTGGATC 1074

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
                | | | ||  ||       |||||| |  |   ||| |||||||| |||||||||
Db       1073 TGGTCGTCTTCGCACAGGGTATAGGCGCGCTCCAGCATGTTTTCCAGCTCGCGGACGTTG 1014

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||  | |       |  | | | | || |  | |||| |   | || | |  || |
Db       1013 CCCGGGAAGCGGTAGTTCTTCAGCTTCTCCTGTGCGTCGCCGGTCAGCCTGGCGGCCGGC 954

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | ||  | | |  | | |||||  || |     |||||||    ||      |||| ||
Db        953 AGGCCGGTGTCGCCGGCCAGGCGCTTGAGGATGCGTTCGGCGAGCAGCGGGATGTCCTCG 894

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              | || || |||||||| | |  || | ||| |  ||     || ||| | || ||||||
Db        893 CGGCGTTCGCGCAGCGG-TGTACGCGCAGCTCGATGACGTTGAGGCGGTAGTAGAGGTCC 835

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||||| |||  ||  ||  ||||  |||| ||| ||| || ||   ||  | |
Db        834 TGGCGGAAGCGCCCGGCGCCGACTTCGGCGGCGAGGTCCTTGTGGGTGGCGCAGAGGATG 775

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGC-CGCACTTCGCCCGTCTGCAA 583
              ||||||| || | ||       |  | | | ||  |||| |||||  |      ||| |
Db        774 CGCACGTGCGACGGCGACCTCCTGCTGGCCGCCGACCGCGCGCACGGCCTTTTCCTGGAT 715

Qy        584 AAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAG 643
                 || |||||     ||||||    ||||||||| ||| |||| |||||||| | |||
Db        714 CGCCCGGAGCAGTTTGACCTGCATGGCCATCGGCAGGTCGGCGACTTCGTCGAGGAACAG 655

Qy        644 CGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGC 703
              || || ||  || |||| | |||| |||  |    | |||  ||||| | |
Db        654 GGTGCCACCGCTGGCGGCCTGGAACAGGCCCTGCTTGTCTTCGATAGCGCCAGTGAAGCT 595

Qy        704 GCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTG 763
              |||   || ||  ||||| | |||||  || ||| || |   ||| |||||  |||| |
Db        594 GCCTTTCTTGTGGCCGAAGAACTCGCTTTCCATCAGCTCGGAGGGAATCGCGCCGCAGTT 535

Qy        764 CACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGT 823
              |||  |||||||||| |||      || || | | ||  ||| | |  |||||||| |
Db        534 CACCGGCACGAACGGCCGCTCGATACGTGGCCCCTGCTCGTGGATCAGGCGCGCCACCAG 475

Qy        824 CTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTT 883
              ||||||||  ||||   ||||||      |  || ||||| |||  ||||||  |||
Db        474 TTCCTTGCCGCTGCCGGACTCGCCACTGATGTAGACCGGCGCCTGGCTGCGCGCCAGCTT 415

Qy        884 GCGCACCTG 892
              ||  | |||
Db        414 GCCGATCTG 406


RESULT 11
US-09-975-719-350/c
; Sequence 350, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
;  APPLICANT: Ausubel, Frederick M.
;  APPLICANT:  Rahme, Laurence G.
;  TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE: 00786/361003
```

```
;  CURRENT APPLICATION NUMBER: US/09/975,719
;  CURRENT FILING DATE:  2001-10-10
;  PRIOR APPLICATION NUMBER: US 09/199,637
;  PRIOR FILING DATE: 1998-11-25
;  PRIOR APPLICATION NUMBER: US 60/066,517
;  PRIOR FILING DATE: 1997-11-25
;  NUMBER OF SEQ ID NOS: 437
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 350
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-975-719-350

  Query Match             24.6%;  Score 221.8;  DB 3;  Length 1344;
  Best Local Similarity   58.3%;  Pred. No. 2.6e-39;
  Matches  425;  Conservative    0;  Mismatches  302;  Indels    2;  Gaps    2;

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              | || | |   | |   |  | ||||  ||| | || |  |||| |     || ||||
Db       1089 GCGGCGCCTTCCTGGCTGGCACCCGGCGCATCGGCCAGGCGCAGGTCGTGAGGCTGGATC 1030

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |  | |  ||  ||       |||||| |  |    ||| |||||||| |||||||||
Db       1029 TGGTCGTCTTCGCACAGGGTATAGGCGCGCTCCAGCATGTTTTCCAGCTCGCGGACGTTG 970

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||  |  |      |  |    | |  | || ||  | | ||||  |   | || |
Db        969 CCCGGGAAGCGGTAGTTCTTCAGCTTCTCCTGTGCGTCGCCGGTCAGCCTGGCGGCCGGC 910

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | ||  |  | |  | | | ||||| || |     |||||||   ||     |||| ||
Db        909 AGGCCGGTGTCGCCGGCCAGGCGCTTGAGGATGCGTTCGGCGAGCAGCGGGATGTCCTCG 850

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              | ||| || |||||||||  | |  || | ||| |  ||    || ||| | || |||||
Db        849 CGGCGTTCGCGCAGCGG-TGTACGCGCAGCTCGATGACGTTGAGGCGGTAGTAGAGGTCC 791

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||||  |||  |||  ||    ||  |||| |||  |||  || ||   || |
Db        790 TGGCGGAAGCGCCCGGCGCCGACTTCGGCGGCGAGGTCCTTGTGGGTGGCGCAGAGGATG 731

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGC-CGCACTTCGCCCGTCTGCAA 583
              ||||||| || | ||    |  | | | ||  |||| ||||  |     |||| |
Db        730 CGCACGTGCGACGGCGACCTCCTGCTGGCCGCCGACCGCGCGCACGGCCTTTTCCTGGAT 671

Qy        584 AAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAG 643
              || |||||     ||||||  |||||||| ||| |||| |||||||| | |||
Db        670 CGCCCGGAGCAGTTTGACCTGCATGGCCATCGGCAGGTCGGCGACTTCGTCGAGGAACAG 611

Qy        644 CGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGC 703
              || || ||    || |||| | |||| |||  |    |  |||  ||||| | |
Db        610 GGTGCCACCGCTGGCGGCCTGGAACAGGCCCTGCTTGTCTTCGATAGCGCCAGTGAAGCT 551

Qy        704 GCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTG 763
              |||   || ||  ||||| | |||||  || ||| || |   ||| |||||  |||| |
Db        550 GCCTTTCTTGTGGCCGAAGAACTCGCTTTCCATCAGCTCGGAGGGAATCGCGCCGCAGTT 491

Qy        764 CACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGT 823
              |||  ||||||||||| |||     || || | | ||  ||| | |  |||||||| |
Db        490 CACCGGCACGAACGGCCGCTCGATACGTGGCCCCTGCTCGTGGATCAGGCGCGCCACCAG 431

Qy        824 CTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTT 883
              ||||||||  ||||  ||||||      |     | ||  ||||| |||  |||||| |||
Db        430 TTCCTTGCCGCTGCCGGACTCGCCACTGATGTGAGACCGGCGCCTGGCTGCGCGCCAGCTT 371

Qy        884 GCGCACCTG 892
              ||  | |||
Db        370 GCCGATCTG 362


RESULT 12
US-09-975-719-1/c
```

```
; Sequence 1, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
;  APPLICANT: Ausubel, Frederick M.
;  APPLICANT:  Rahme, Laurence G.
;  TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE: 00786/361003
;  CURRENT APPLICATION NUMBER: US/09/975,719
;  CURRENT FILING DATE:  2001-10-10
;  PRIOR APPLICATION NUMBER: US 09/199,637
;  PRIOR FILING DATE: 1998-11-25
;  PRIOR APPLICATION NUMBER: US 60/066,517
;  PRIOR FILING DATE: 1997-11-25
;  NUMBER OF SEQ ID NOS: 437
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 42235
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(42235)
;   OTHER INFORMATION: N is any nucleic acid.
US-09-975-719-1

  Query Match             24.6%;  Score 221.8;  DB 3;  Length 42235;
  Best Local Similarity   58.3%;  Pred. No. 1.3e-39;
  Matches  425;  Conservative    0;  Mismatches  302;  Indels    2;  Gaps    2;

Qy        165 GAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATC 224
              | || | |   | |   |  | |||||  ||| | || |  |||| |     || ||||
Db      32354 GCGGCGCCTTCCTGGCTGGCACCCGGCGCATCGGCCAGGCGCAGGTCGTGAGGCTGGATC 32295

Qy        225 GCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTG 284
              |  | |  ||  ||       |||||| |  |    ||| |||||||| |||||||||
Db      32294 TGGTCGTCTTCGCACAGGGTATAGGCGCGCTCCAGCATGTTTTCCAGCTCGCGGACGTTG 32235

Qy        285 CCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCC 344
              |||||  |  |      |  |    | |  | || ||  | | ||||  |   | || |
Db      32234 CCCGGGAAGCGGTAGTTCTTCAGCTTCTCCTGTGCGTCGCCGGTCAGCCTGGCGGCCGGC 32175

Qy        345 TGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG 404
              | ||  |  | |  | | | ||||| || |     |||||||   ||     |||| ||
Db      32174 AGGCCGGTGTCGCCGGCCAGGCGCTTGAGGATGCGTTCGGCGAGCAGCGGGATGTCCTCG 32115

Qy        405 GGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCG 464
              | ||  || ||||||||  | |  || | ||| |  ||    || ||| | || |||||
Db      32114 CGGCGTTCGCGCAGCGG-TGTACGCGCAGCTCGATGACGTTGAGGCGGTAGTAGAGGTCC 32056

Qy        465 CTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGC 524
              |||||||  |||  |||  ||    ||  |||| |||  |||  || ||   || |
Db      32055 TGGCGGAAGCGCCCGGCGCCGACTTCGGCGGCGAGGTCCTTGTGGGTGGCGCAGAGGATG 31996

Qy        525 CGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGC-CGCACTTCGCCCGTCTGCAA 583
              ||||||| || | ||    |  | | | ||  |||| ||||  |     |||| |
Db      31995 CGCACGTGCGACGGCGACCTCCTGCTGGCCGCCGACCGCGCGCACGGCCTTTTCCTGGAT 31936

Qy        584 AAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAG 643
              || |||||     ||||||  |||||||| ||| |||| |||||||| | |||
Db      31935 CGCCCGGAGCAGTTTGACCTGCATGGCCATCGGCAGGTCGGCGACTTCGTCGAGGAACAG 31876

Qy        644 CGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGC 703
              || || ||    || |||| | |||| |||  |    |  |||  ||||| | |
Db      31875 GGTGCCACCGCTGGCGGCCTGGAACAGGCCCTGCTTGTCTTCGATAGCGCCAGTGAAGCT 31816

Qy        704 GCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTG 763
              |||   || ||  ||||| | |||||  || ||| || |   ||| |||||  |||| |
Db      31815 GCCTTTCTTGTGGCCGAAGAACTCGCTTTCCATCAGCTCGGAGGGAATCGCGCCGCAGTT 31756

Qy        764 CACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGT 823
              |||  ||||||||||| |||     || || | | ||  ||| | |  |||||||| |
Db      31755 CACCGGCACGAACGGCCGCTCGATACGTGGCCCCTGCTCGTGGATCAGGCGCGCCACCAG 31696
```

```
Qy        824 CTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTT 883
              ||||||||  ||||   ||||||     |      | || ||||| |||  |||||| |||
Db      31695 TTCCTTGCCGCTGCCGGACTCGCCACTGATGTAGACCGGCGCCTGGCTGCGCGCCAGCTT 31636

Qy        884 GCGCACCTG 892
              ||  | |||
Db      31635 GCCGATCTG 31627


RESULT 13
US-10-411-910A-273/c
; Sequence 273, Application US/10411910A
; Publication No. US20040209256A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Harrison F.
;  TITLE OF INVENTION: Methods and Compositions for Evolving Hydrogenase Genes
;  FILE REFERENCE: H2041203-P
;  CURRENT APPLICATION NUMBER: US/10/411,910A
;  CURRENT FILING DATE:  2003-04-12
;  NUMBER OF SEQ ID NOS: 343
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 273
;   LENGTH: 2373
;   TYPE: DNA
;   ORGANISM: Geobaccer metallireducens
US-10-411-910A-273

  Query Match             24.6%;  Score 221.2;  DB 8;  Length 2373;
  Best Local Similarity   56.2%;  Pred. No. 3.1e-39;
  Matches  415;  Conservative    0;  Mismatches  323;  Indels    0;  Gaps    0;

Qy        163 CGGAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGA 222
              ||| || || |  |     | | | |     | | ||   || ||   ||      ||
Db       2176 CGGTGGCGAAGCCCTCCTCCAGGCCCTCGCTCAGGGCCAGGGGGAAGTTCTCCACGCGGA 2117

Qy        223 TCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGT 282
              || ||| | |    ||  |||||  |||||||  |   |    |||||| || ||| |||| ||
Db       2116 TCACGCCGTCGTGGGTCAGCACGGCGGCGCGCTCGATCACGTTCTGCATCTCGCGGATGT 2057

Qy        283 TGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCG 342
              |||| |||||  ||   |  |  | ||||      ||    ||   |  | | |
Db       2056 TGCCGGGCCAGGGGTACTTGGTCAGGGCCTTCACGGCGCGGTCCTCGATGCCCACCATGC 1997

Qy        343 CCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCC 402
              ||    || |      |      | ||  |      |   | | |       | |
Db       1996 GCTTGTTCAGCTTGGTGCTGGCCTTGTCCATGAAGTGGTGCACCAGGATGGGGATGCTCT 1937

Qy        403 CGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGT 462
              |    ||     ||||| || ||||    |||     | |||   ||| ||| | |||||||
Db       1936 CCACGCGGCTGCGCAGGGGGGGCATGGTGATGGTGAACACGTTCAGGCGGTAGTACAGGT 1877

Qy        463 CGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCA 522
              |    |||||  |  |||||||||||    |  | | || | |  ||  || || |||||||
Db       1876 CCTCGCGGAACCAGCCCTCGCGCACGCCCTGCTCCAGGTTCTTGTTGGTGGCGGCGATCA 1817

Qy        523 GCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCA 582
              | |||||||||| ||  |  ||||   ||     |||||  | ||||| ||||||  | | ||
Db       1816 GGCGCACGTCCACCTTCTTGGCGGCGGTGCCGCCCACGGGGCGCACCTCGCCCATGTCCA 1757

Qy        583 AAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCA 642
              |  ||||||||     ||||||  |||     |   |  |||||| |||||  || |  |
Db       1756 GCACGCGCAGCAGCTCGGCCTGCAGCTTGGGGGTGATGTCGCCGATCTCGTCCAGGAAGA 1697

Qy        643 GCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCG 702
              || ||||| || || |||||| |||| |||  |    | ||  |  |||||  | |   |
Db       1696 TGGTGCCGCCGTCGGCGGCCTCGAACAGGCCCTTCTTGTCGGCCACGGCGCCGGTGAAGG 1637

Qy        703 CGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACT 762
              |||||  || ||  |||||||||||||| ||| | | |     |||| |||  |  | ||| |
Db       1636 CGCCGCGCTTGTGGCCGAACAGCTCGCTCTCCAGCAGGCTGTCGGTGATGGTGGTGCAGT 1577

Qy        763 GCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCG 822
```

```
              ||||  |||  |  ||||  |     |||  |||    |||||  |  |||||  ||   |
Db       1576 TCACGGTCACCAGGGGCTTGTCGTTGCGCTTGCTGTAGCGGTGGATGGCGCGGGCGGTCA 1517

Qy        823 TCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCT 882
              |||||||||  |||||  | ||||||     | |||||  || |     | | ||  ||||||
Db       1516 GCTCCTTGCCGGTGCCGCTCTCGCCGCGGATCAGCACGGTGGTGGGGGTGGGGCCCACCT 1457

Qy        883 TGCGCACCTGCGCGAACA 900
                | | |  | | ||
Db       1456 GCTTGATCAGGCCCAGCA 1439


RESULT 14
US-10-282-122A-11630/c
; Sequence 11630, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11630
;   LENGTH: 1359
;   TYPE: DNA
;   ORGANISM: Burkholderia cepacia
US-10-282-122A-11630

  Query Match             24.4%;  Score 219.4;  DB 7;  Length 1359;
  Best Local Similarity   59.5%;  Pred. No. 8.6e-39;
  Matches  371;  Conservative    0;  Mismatches  253;  Indels    0;  Gaps    0;

Qy        243 ACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCC 302
              | ||||  |||||    |  |||||| |||||| || |||||||| |||||||    || |
Db       1124 AGCACGAAGCGGTCGGCCACGTTCTGCAGCTCGCGCACGTTGCCGGGCCACGCGTGCGTC 1065

Qy        303 GCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTC 362
              |  | |  |  | | || |  |  ||| || |||||| |    |||||
Db       1064 ATCAGTTCGGATACCTGCGCGGCGGTGACGACCGGCGCCGGCTGCCCGAAGCGCCGCGCG 1005
```

```
Qy        363 CGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGC 422
              ||  |    ||||    |   ||  | ||||| || |||| || ||||||||||
Db       1004 GCCGCGAGCACGAAATGCTCGAACAGCAGCGGCACGTCCTCGCGCCGTTCGCGCAGCGGC 945

Qy        423 GGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCG 482
              ||||||  ||||| |   ||    || ||  |   ||| ||    || || |  || |||
Db        944 GGCAGCTCGATCTGCGCGACGTTGAGGCGATACAGCAGATCCGCGCGAAAACGTCCGTCG 885

Qy        483 CGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGG 542
              | | |   |  ||||||||  | || ||| || | ||  ||   |||  || |
Db        884 GCCGCGAGTTCGGCGAGATCCGCCTTCGACGCGGCCACCACGCGGCAGTCGACCGGAATC 825

Qy        543 GCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCC 602
              |   |||  || | ||||   ||  |||    |||||  |  |||||||          |
Db        824 AGGTCGTTCGCGCCGAGCCGCTCGACCACGCGTTCCTGCAGCACGCGCAGCATCTTGATC 765

Qy        603 TGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCC 662
              ||||  |   ||||||| |    ||| |||||||||  | |||||| |||||  ||||||  |
Db        764 TGCAGCGCGATCGGCATCGTCTCGATCTCGTCGAGGAACAGCGTGCCGCCGTCCGCCCAC 705

Qy        663 TCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAAC 722
              |||  |   ||   |||  |  || ||||||  | |  ||||||||||||||  ||||||
Db        704 TCGATCTTGCCGATGCGCTTCTTGATCGCGCCGGTGAACGCGCCGGCCTCGTGGCCGAAC 645

Qy        723 AGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGC 782
              |||||||  |||||  |   |  |||||  |    |  |||| | ||     |||||  ||
Db        644 AGCTCGCTCTCGAACACCTGCTCGGGCAGGCCGCCGCAGTTCAGCGCGACGAAATGCGCG 585

Qy        783 TGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCC 842
              |  || ||   ||  |    |||||    ||||||  | |  |||||||||  ||||||  |
Db        584 TCGCGGCGGCCGCCGAAGTCGTGCAGGCTGCGCGCGATCAACTCCTTGCCGGTGCCCGTC 525

Qy        843 TCGCCCACCACCAGCAGCGTCGCC 866
              |||||    | ||||| || | ||
Db        524 TCGCCGGTGATCAGCACCGACACC 501


RESULT 15
US-09-870-162A-22
; Sequence 22, Application US/09870162A
; Patent No. US20020042118A1
; GENERAL INFORMATION:
;  APPLICANT: Breinig, Sabine
;  APPLICANT:  Fuchs, Georg
;  TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica
;  FILE REFERENCE: BC1006 US DIV
;  CURRENT APPLICATION NUMBER: US/09/870,162A
;  CURRENT FILING DATE:  2001-05-30
;  PRIOR APPLICATION NUMBER: 09/516914
;  PRIOR FILING DATE: 2000-03-01
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 1749
;   TYPE: DNA
;   ORGANISM: Thauera aromatica
US-09-870-162A-22

  Query Match             24.3%;  Score 218.6;  DB 3;  Length 1749;
  Best Local Similarity   58.0%;  Pred. No. 1.2e-38;
  Matches  386;  Conservative    0;  Mismatches  279;  Indels    0;  Gaps    0;

Qy        226 CGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGC 285
              ||| |  |   || || | |||  |||| |  || | ||||| ||| || |||| |||||
Db        345 CGCCGTTCGGCGCAAGGATCACCCCGCGTTCCACCGTGTTCTGCAGTTCGCGGATGTTGC 404

Qy        286 CCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCT 345
              |||||||  |      |  | |  |||     |||||   | || |   ||||| ||  ||
Db        405 CCGGCCAGGCATGGCCGACCAGCGCCTTCTTCGCCTTGTCGGAAAATCCGAGCAGCTTCT 464

Qy        346 GCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGG 405
                ||  | | |  ||    |  | | || |    || ||| |   |      |||| |

Db        465 TGCCGTTGATCGCCGCATATTTCCTGAGGAAATGCTTGGCCAGAGGGGAGATGTCCTCCC 524

Qy        406 GCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGC 465
              ||||  |||||||||||| |    |||||  |   |    || ||| | |||||||||
Db        525 TGCGCTGGCGCAGCGGCGGAATGTCGATCTGGAAAGCATTGAGACGGTAGTACAGGTCGG 584

Qy        466 TCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCC 525
              || ||| | ||| ||   ||||| |    ||||         |  ||||| ||||  || |
Db        585 CCCTGAAACGCCCTTCCTTCACCAACTGGGCGAGGCTGGCATTGGTCGCGGCGACGAGGC 644

Qy        526 GCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAA 585
              | |||||    || ||||           ||||  |||   || |||||   ||| |  |
Db        645 GGACGTCGATACGGCGGGTCTTGTCATCGCCCAAACGCTCGACCTCGCCTTCCTGGAGCA 704

Qy        586 AACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCG 645
              ||||||||    | |||| || ||| |||||| |||||||  |||||  || | ||| |
Db        705 CCCGCAGCAGCTTGCTCTGCGCCGTCAGCGGCAGATCGCCGATTTCGTCCAGGAACAGGG 764

Qy        646 TCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGC 705
              | ||||| |  ||   |||| ||   ||||  || || || |||||||| | |    || |
Db        765 TGCCGCCGTTGGCGCGCTCGAACCTGCCCGGGCGGGCTGCCAGCGCGCCGGTGTATGCCC 824

Qy        706 CGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCGCACTGCA 765
              |    ||    ||||| |||||| |||| |   |  | |  || ||||| ||||| |  |
Db        825 CTTTTTCCACGCCGAAAAGCTCGGCCTCCACGAGGTCGTGCGGAATCGCGGCGCAGTTGA 884

Qy        766 CGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCT 825
              ||  | |||  | |  || || | |||||    ||||| |||||||||  | |   |
Db        885 TCGCAACCAGCGGGCGATCGCGGCGGGCGCTCATTTCGTGCAGCGCGCGCGTGAACAGTT 944

Qy        826 CCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGC 885
              |||||||    |||  ||||| | || ||  |  |        |||   |||  |||
Db        945 CCTTGCCGACCCCCGTTTCGCCGAGCAGCAAAATGGCGATGCTGCTGCCCGCGGCCTGCT 1004

Qy        886 GCACC 890
              ||| |
Db       1005 GCAGC 1009


Search completed: April 27, 2006, 14:39:34
Job time : 903.315 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 565.152 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_1_900
Perfect score: 900
Sequence: 1 tgggtatcccgagccgctgg.....cttgcgcacctgcgcgaaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	258.2	28.7	1386	14	US-11-075-185-55	Sequence 55, Appl
c	2	258.2	28.7	14172	14	US-11-075-185-2	Sequence 2, Appli
c	3	167.4	18.6	3363	14	US-11-091-914-1	Sequence 1, Appli
c	4	149	16.6	1371	11	US-11-079-463-2998	Sequence 2998, Ap
c	5	140.2	15.6	1359	11	US-11-079-463-1701	Sequence 1701, Ap
c	6	135	15.0	1569	11	US-11-079-463-4388	Sequence 4388, Ap
c	7	132.2	14.7	1542	14	US-11-149-349-5	Sequence 5, Appli
c	8	117	13.0	1404	11	US-11-079-463-4832	Sequence 4832, Ap
c	9	111.8	12.4	1251	11	US-11-079-463-888	Sequence 888, App
c	10	108.8	12.1	1422	11	US-11-079-463-2940	Sequence 2940, Ap
c	11	108	12.0	1017	11	US-11-079-463-3432	Sequence 3432, Ap
c	12	107.8	12.0	1434	11	US-11-079-463-549	Sequence 549, App
c	13	103.2	11.5	1395	11	US-11-079-463-291	Sequence 291, App
	14	101.2	11.2	765	14	US-11-075-185-50	Sequence 50, Appl
	15	101.2	11.2	14172	14	US-11-075-185-2	Sequence 2, Appli
c	16	89.4	9.9	5679	14	US-11-075-185-36	Sequence 36, Appl
c	17	89.4	9.9	78869	14	US-11-075-185-1	Sequence 1, Appli
c	18	89.2	9.9	14055	7	US-10-496-351-1	Sequence 1, Appli
	19	89.2	9.9	82746	7	US-10-496-351-56	Sequence 56, Appl
	20	86.4	9.6	2944528	11	US-11-045-004-1	Sequence 1, Appli
c	21	86.2	9.6	88421	14	US-11-205-109-1	Sequence 1, Appli
c	22	84.8	9.4	5706	14	US-11-052-554A-519	Sequence 519, App
c	23	84.6	9.4	1398	14	US-11-098-686-9587	Sequence 9587, Ap
	24	84.6	9.4	1457619	14	US-11-098-686-8739	Sequence 8739, Ap
c	25	84.4	9.4	1050	14	US-11-098-686-9992	Sequence 9992, Ap
c	26	84.4	9.4	4983	14	US-11-052-554A-521	Sequence 521, App
c	27	84.4	9.4	9222	7	US-10-496-351-2	Sequence 2, Appli
c	28	84.4	9.4	1457619	14	US-11-098-686-8739	Sequence 8739, Ap
c	29	84.2	9.4	82746	7	US-10-496-351-56	Sequence 56, Appl
c	30	81.8	9.1	10968	14	US-11-075-185-35	Sequence 35, Appl
c	31	80.2	8.9	1416	14	US-11-098-686-9979	Sequence 9979, Ap
	32	80.2	8.9	2183	11	US-11-072-512-1064	Sequence 1064, Ap
c	33	80	8.9	2562	14	US-11-052-554A-533	Sequence 533, App
c	34	80	8.9	11070	14	US-11-075-185-34	Sequence 34, Appl
c	35	79	8.8	1722	8	US-10-467-657-2547	Sequence 2547, Ap
	36	78	8.7	30191	9	US-10-330-773-631	Sequence 631, App
c	37	77.8	8.6	11905	7	US-10-496-351-57	Sequence 57, Appl
c	38	76.4	8.5	885	7	US-10-496-351-14	Sequence 14, Appl
c	39	76.4	8.5	10524	14	US-11-075-185-37	Sequence 37, Appl
	40	75	8.3	3354	7	US-10-496-351-12	Sequence 12, Appl
c	41	74.8	8.3	3240	14	US-11-052-554A-529	Sequence 529, App
c	42	74.4	8.3	9975	7	US-10-496-351-4	Sequence 4, Appli
c	43	74.2	8.2	1374	14	US-11-098-686-9777	Sequence 9777, Ap
c	44	74	8.2	1314	7	US-10-496-351-42	Sequence 42, Appl
c	45	73.8	8.2	1848	14	US-11-052-554A-536	Sequence 536, App

ALIGNMENTS

RESULT 1
US-11-075-185-55/c
; Sequence 55, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 55
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-55

Query Match 28.7%; Score 258.2; DB 14; Length 1386;
Best Local Similarity 60.7%; Pred. No. 1e-40;
Matches 422; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

```
Qy        206 GAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTT 265
              ||| | | |       ||||     |  ||  |||||| | || || || | |  |||  |||
Db       1152 GAGATCGTCGACGGTGATCTGATCGTACCGCGCGAGCGCGACCGCCCGCTCGACGCAGTT 1093

Qy        266 CTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGC 325
              ||  ||||| ||||||||||| |||||||     ||  |||     | ||  |||   |||
Db       1092 CTGGAGCTCGCGGACGTTGCCTGGCCACGCGTAGGCGACGAGCTTCTCGGCCGCGGGCGC 1033

Qy        326 CGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGC 385
              ||| ||  |       || | || || | ||   ||       |||  || |    ||  ||
Db       1032 CGAGAGGCCCTTGATCGGCCGCTCCTTGACGGTCGCGAAGTGCTCGAGGAAGCGCTGCGC 973

Qy        386 GATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCG 445
```

```
Db        972 GAGCAGGAGCACGTCGGCGCCGCGGGAGCGGAGCGGCGGCAGATCGACGTGGACCACGTT 913

Qy        446 CAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCG 505

Db        912 GATCCGGTAATAGAGGTCCTCGCGGAAGCGGCGCTCCTCGATCGCGGTCTCCAGGTCGCG 853

Qy        506 GTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCG 565

Db        852 GTTCGTCGCCGCGATGAGCCGCACGTCCACGGGGACCTCCTCGTCCGCGCCGACGGGGCG 793

Qy        566 CACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCC 625

Db        792 GACGACGCGCTCCTGGAGGGCGCGCAGGAGCTTCGGCTGGAGCCCGACCGGCAGCTCGCC 733

Qy        626 GACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGC 685

Db        732 GATCTCGTCGAGGAACAGGGTGCCGCCGTGGGCCCGCGCGAACAGGCCGCTCCTCGGCCC 673

Qy        686 GAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATC 745

Db        672 CTTGGCGTCGGTGAAGGCGCCCCGCGCGTGGCCGAACAGCTCGCTCTCGAGCAGGGCGTC 613

Qy        746 GGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTG 805

Db        612 CGGGACCGCCGCGCAGTTCACCGCGACGAACGCGCCCTGGCCGCGCCGGCTGCGCTGGTG 553

Qy        806 CACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGC 865

Db        552 CACGGCGCGCGCGACGAGCTCCTTGCCGGTGCCGCTCTCGCCCGTGATGAGGATCGACGT 493

Qy        866 CTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900

Db        492 GTCCGAGTCGGCGACCCGGTCGAGCAGATCGAAGA 458


RESULT 2
US-11-075-185-2/c
; Sequence 2, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
;   LENGTH: 14172
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-2

  Query Match             28.7%;  Score 258.2;  DB 14;  Length 14172;
  Best Local Similarity   60.7%;  Pred. No. 6e-41;
  Matches  422;  Conservative    0;  Mismatches  273;  Indels    0;  Gaps    0;

Qy        206 GAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTT 265

Db       8430 GAGATCGTCGACGGTGATCTGATCGTACCGCGCGAGCGCGACCGCCCGCTCGACGCAGTT 8371

Qy        266 CTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGC 325

Db       8370 CTGGAGCTCGCGGACGTTGCCTGGCCACGCGTAGGCGACGAGCTTCTCGGCCGCGGGCGC 8311

Qy        326 CGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGC 385

Db       8310 CGAGAGGCCCTTGATCGGCCGCTCCTTGACGGTCGCGAAGTGCTCGAGGAAGCGCTGCGC 8251

Qy        386 GATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCG 445

Db       8250 GAGCAGGAGCACGTCGGCGCCGCGGGAGCGGAGCGGCGGCAGATCGACGTGGACCACGTT 8191

Qy        446 CAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCG 505

Db       8190 GATCCGGTAATAGAGGTCCTCGCGGAAGCGGCGCTCCTCGATCGCGGTCTCCAGGTCGCG 8131

Qy        506 GTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCG 565

Db       8130 GTTCGTCGCCGCGATGAGCCGCACGTCCACGGGGACCTCCTCGTCCGCGCCGACGGGGCG 8071

Qy        566 CACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCC 625

Db       8070 GACGACGCGCTCCTGGAGGGCGCGCAGGAGCTTCGGCTGGAGCCCGACCGGCAGCTCGCC 8011

Qy        626 GACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGC 685

Db       8010 GATCTCGTCGAGGAACAGGGTGCCGCCGTGGGCCCGCGCGAACAGGCCGCTCCTCGGCCC 7951

Qy        686 GAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATC 745

Db       7950 CTTGGCGTCGGTGAAGGCGCCCCGCGCGTGGCCGAACAGCTCGCTCTCGAGCAGGGCGTC 7891

Qy        746 GGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTG 805

Db       7890 CGGGACCGCCGCGCAGTTCACCGCGACGAACGCGCCCTGGCCGCGCCGGCTGCGCTGGTG 7831

Qy        806 CACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGC 865

Db       7830 CACGGCGCGCGCGACGAGCTCCTTGCCGGTGCCGCTCTCGCCCGTGATGAGGATCGACGT 7771

Qy        866 CTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900

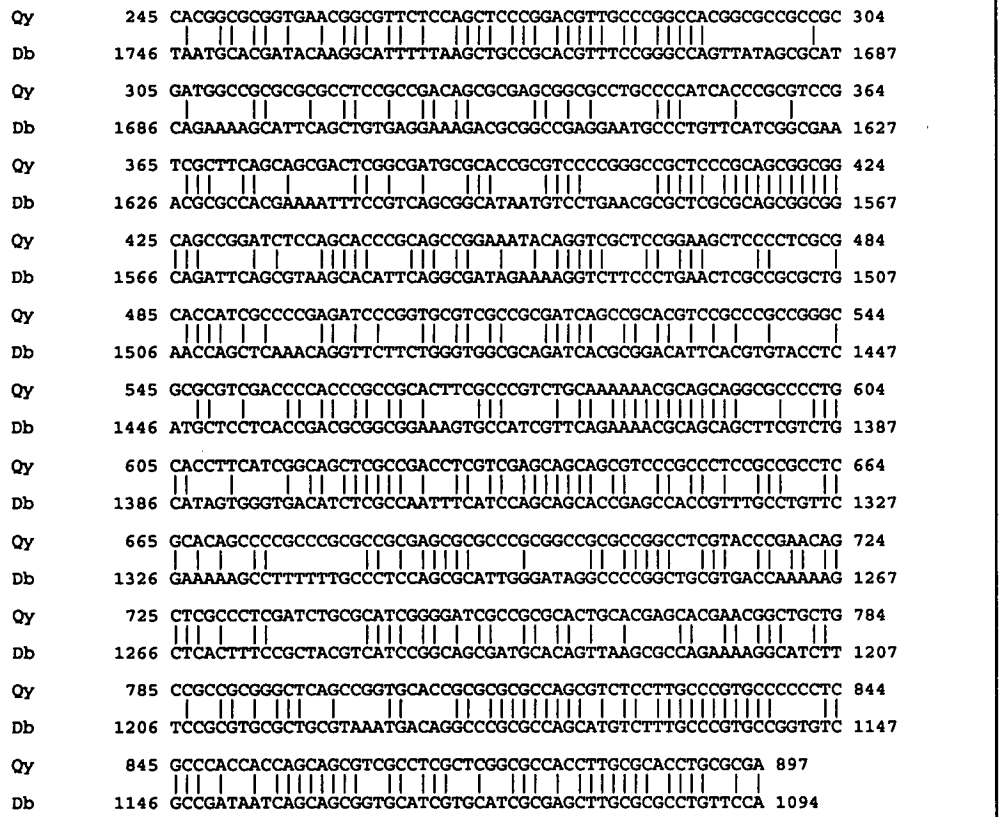
Db       7770 GTCCGAGTCGGCGACCCGGTCGAGCAGATCGAAGA 7736


RESULT 3
US-11-091-914-1/c
; Sequence 1, Application US/11091914
; Publication No. US20050277174A1
; GENERAL INFORMATION:
;  APPLICANT: Hidehiko KUMAGAI, Hideyuki SUZUKI, Takane KATAYAMA, Miyoko NAWATA and
;  APPLICANT:  Hidetsugu NAKAZAWA
;  TITLE OF INVENTION: MUTANT TYROSINE REPRESSOR, A GENE ENCODING THE SAME, AND A METHOD
;  TITLE OF INVENTION:  PRODUCING L-DOPA
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/11/091,914
;  CURRENT FILING DATE:  2005-03-29
;  PRIOR APPLICATION NUMBER: JP 2004-95147
;  PRIOR FILING DATE: 2004-03-29
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 3363
;   TYPE: DNA
;   ORGANISM: Pantoea agglomerans
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (442)..(2004)
;   OTHER INFORMATION:
US-11-091-914-1

  Query Match             18.6%;  Score 167.4;  DB 14;  Length 3363;
  Best Local Similarity   52.2%;  Pred. No. 1.6e-23;
  Matches  372;  Conservative    0;  Mismatches  341;  Indels    0;  Gaps    0;

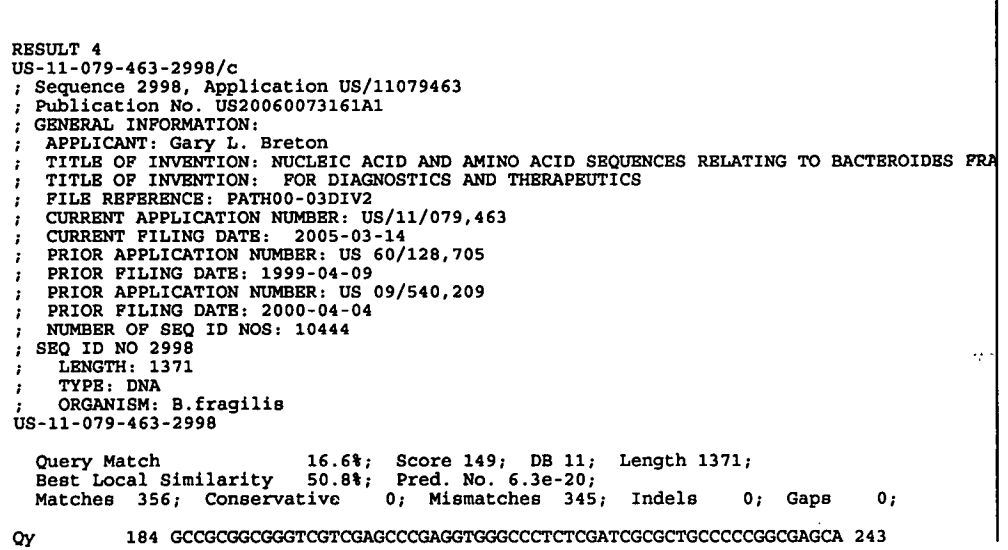
Qy        185 CCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCAC 244

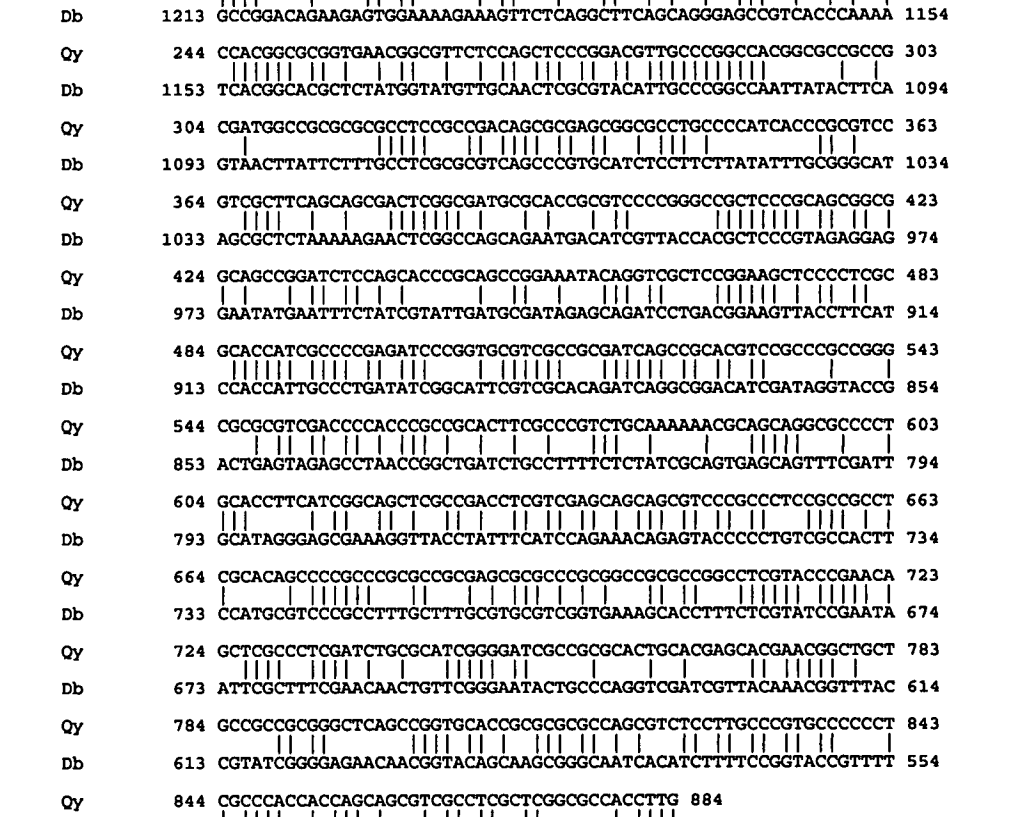
Db       1806 CAGCGCCTGTTCCGGCAAGACGATATCCTGCGGCCGTAACTCATGGCCTTCCAACTGGGT 1747
```



RESULT 4
US-11-079-463-2998/c
; Sequence 2998, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 2998
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-2998

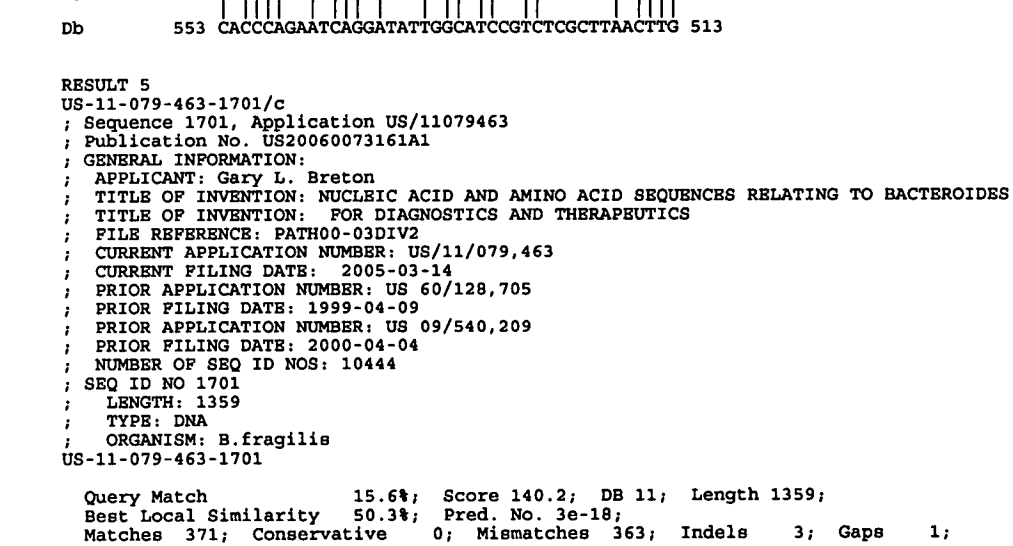
Query Match 16.6%; Score 149; DB 11; Length 1371;
Best Local Similarity 50.8%; Pred. No. 6.3e-20;
Matches 356; Conservative 0; Mismatches 345; Indels 0; Gaps 0;

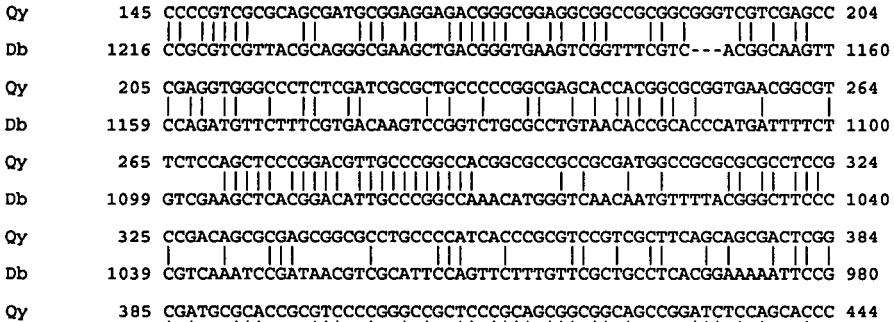




RESULT 5
US-11-079-463-1701/c
; Sequence 1701, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1701
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-1701

Query Match 15.6%; Score 140.2; DB 11; Length 1359;
Best Local Similarity 50.3%; Pred. No. 3e-18;
Matches 371; Conservative 0; Mismatches 363; Indels 3; Gaps 1;





RESULT 6
US-11-079-463-4388/c
; Sequence 4388, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 4388
; LENGTH: 1569
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-4388

Query Match 15.0%; Score 135; DB 11; Length 1569;
Best Local Similarity 48.7%; Pred. No. 2.8e-17;
Matches 366; Conservative 0; Mismatches 385; Indels 0; Gaps 0;

RESULT 7
US-11-149-349-5/c
; Sequence 5, Application US/11149349
; Publication No. US20050277179A1
; GENERAL INFORMATION:
; APPLICANT: Atsuko TAKAI, Ranko NISHI, Yuji JOE, Hisao ITO
; TITLE OF INVENTION: L-tyrosine-producing bacterium and a method for producing L-tyros
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/149,349
; CURRENT FILING DATE: 2005-06-10
; PRIOR APPLICATION NUMBER: JP2004-176797
; PRIOR FILING DATE: 2004-06-15
; PRIOR APPLICATION NUMBER: JP2005-112484

```
;  PRIOR FILING DATE: 2005-04-08
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 1542
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1542)
;   OTHER INFORMATION:
US-11-149-349-5

  Query Match            14.7%;  Score 132.2;  DB 14;  Length 1542;
  Best Local Similarity  51.9%;  Pred. No. 9.7e-17;
  Matches  339;  Conservative    0;  Mismatches  293;  Indels   21;  Gaps    1;

Qy        245 CACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGC 304
Db       1281 CAGTGCGCGATAGATAGCGTTCTTTAACTGCCGCACATTTCCCGGCCACGCATAACGCGT 1222

Qy        305 GATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCG 364
Db       1221 AAGTACAGTATTCAGGTCAGCGGCCAGTTTCGGACGCGGCACGCCCTGCTCGTCGGCAAA 1162

Qy        365 TCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGG 424
Db       1161 GCGGGCGACGAACAGCTCAGTTAACGGCATGATGTCCTGCGGACAGTCACGTAGCGGCGG 1102

Qy        425 CAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCG 484
Db       1101 CAGATTGAGCGTCAACACGTTCAGACGATAATAGAGATCTTCACGGAACATGCCTTTTTG 1042

Qy        485 CACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGC 544
Db       1041 CACCAGTTCGACCAGATTCTTCTGCGTAGCGCAAATCACCCGCACATCGACATGCACCTC 982

Qy        545 GCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTG 604
Db        981 ATGGTCTTCGCCAACCCGACGGAAAGTGCCATCATTAAGGAAACGCAGTAATTTCGCCTG 922

Qy        605 CACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTC 664
Db        921 CATCCGTGGTGACATTTCCCCTATTTCATCCAACAGCACCGAACCACCGTTCGCCTGCTC 862

Qy        665 GCACAGCCCCGCCCGCCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAG 724
Db        861 AAAGAATCCTTTCTTCCCTTCCGGAGCA----------TGACCAAACAG 823

Qy        725 CTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTG 784
Db        822 TTCACTCTCGACCGCATCTTCCGGTATAGACGCACAGTTCAGCGCCAGGTAAGGTTTGCC 763

Qy        785 CCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTC 844
Db        762 CGCTCTGGGGCTTGCCTGATGGCAGGCGTAGGCAAAGAGATCTTTACCTGTACCTGTGTC 703

Qy        845 GCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGA 897
Db        702 ACCCGTAATCAGCAGCGGCGCGCTTAGCATCGCCAGTTTCTGCGCCTGTTCGA 650


RESULT 8
US-11-079-463-4832/c
; Sequence 4832, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH00-03DIV2
;  CURRENT APPLICATION NUMBER: US/11/079,463
;  CURRENT FILING DATE:  2005-03-14
;  PRIOR APPLICATION NUMBER: US 60/128,705
;  PRIOR FILING DATE: 1999-04-09
;  PRIOR APPLICATION NUMBER: US 09/540,209
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 4832
;   LENGTH: 1404
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-4832

  Query Match            13.0%;  Score 117;  DB 11;  Length 1404;
  Best Local Similarity  51.4%;  Pred. No. 7.8e-14;
  Matches  270;  Conservative    0;  Mismatches  255;  Indels    0;  Gaps    0;

Qy        375 AGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATC 434
Db       1040 AACGATTCAGCCAGTATCTTAATATCTCCCGTACGTTCACGCAAAGGCGGAAGATGTACC 981

Qy        435 TCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCC 494
Db        980 TGAAACACGGAGAGCCGATAGAACAGATCTTCACGAAAACGTCCCGCAGCTATCTCGGCA 921

Qy        495 CCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGAC 554
Db        920 GGTAAATTACGGTTTGTAGCAGCAATGATACGCACATCAATGTGGGTCGGCTTCGTGTCA 861

Qy        555 CCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATC 614
Db        860 CCTATCTTGATGTATTCGCCGGTTTCGAGGATACGCAACAGCTTGGCTTGCAACTCAAAA 801

Qy        615 GGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCC 674
Db        800 GCCATCTCACCGATTTCATCCAAAAAGATTGTTCCGTTATTGGCTTCTTCAAAAAGTCCC 741

Qy        675 GCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCG 734
Db        740 TTCTTATCTTTCAACGCACCGGTAAAGGAACCCGCCTTATGCCCGAACATCTCGCTCTCC 681

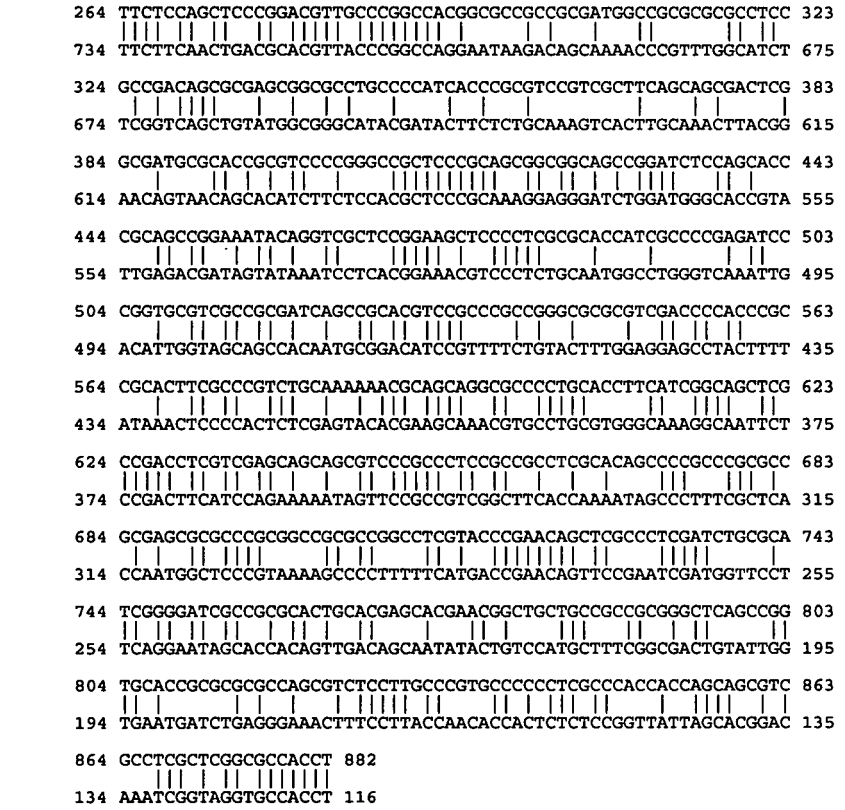
Qy        735 ATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGG 794
Db        680 AGCAACTCTTTACTAAACGAAGAGCAGTTAACTGCCACAAAGTTTTGCCGGGCACGCTTA 621

Qy        795 CTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACC 854
Db        620 CTACTATAATGTATGGACTGGGCAAACACCTCTTTTCCGGTTCCCGTCTCACCCGTCAGC 561

Qy        855 AGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAAC 899
Db        560 AATACAGGTACATCAGTCACCGATACCTTACGGGCTAACGAAACC 516

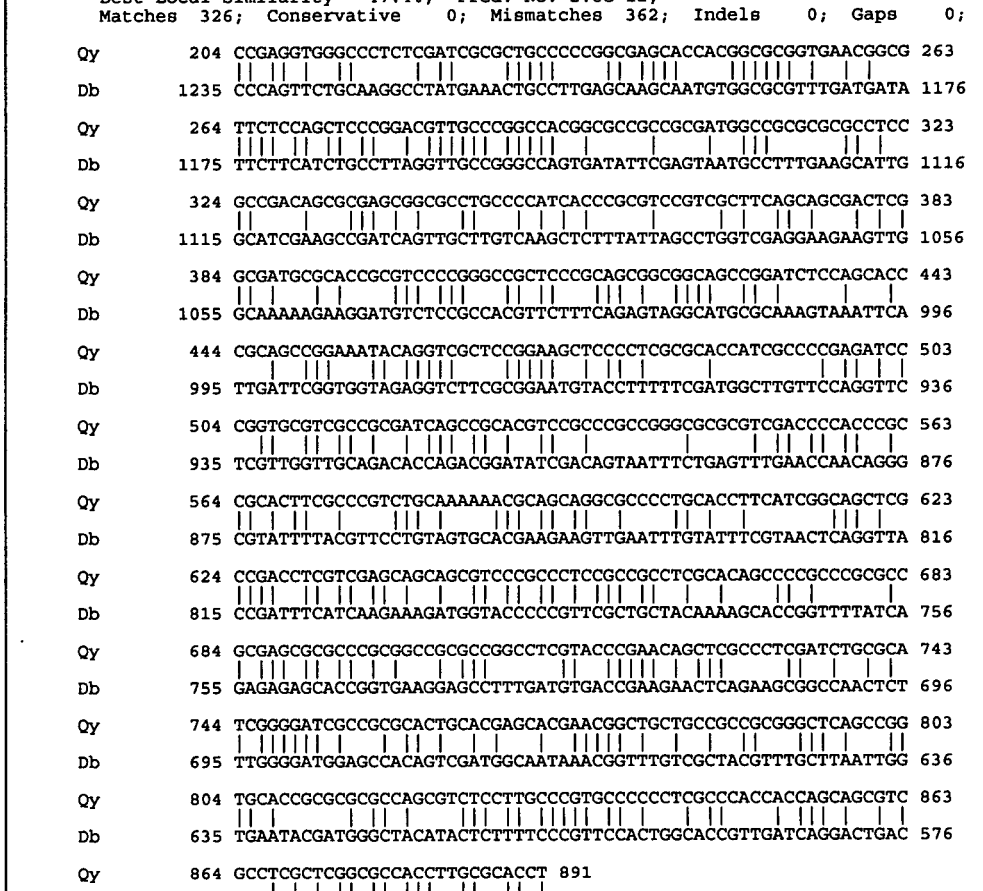

RESULT 9
US-11-079-463-888/c
; Sequence 888, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH00-03DIV2
;  CURRENT APPLICATION NUMBER: US/11/079,463
;  CURRENT FILING DATE:  2005-03-14
;  PRIOR APPLICATION NUMBER: US 60/128,705
;  PRIOR FILING DATE: 1999-04-09
;  PRIOR APPLICATION NUMBER: US 09/540,209
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 888
;   LENGTH: 1251
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-888
```

Query Match 12.4%; Score 111.8; DB 11; Length 1251;
Best Local Similarity 48.8%; Pred. No. 7.8e-13;
Matches 302; Conservative 0; Mismatches 317; Indels 0; Gaps 0;



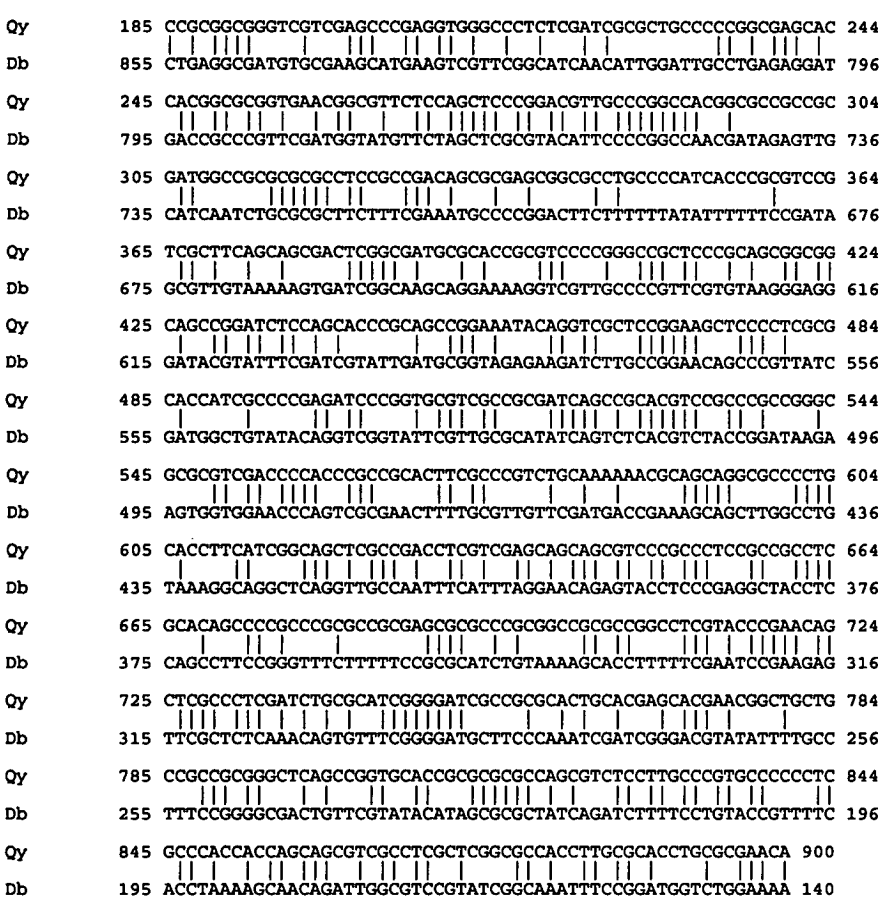
RESULT 10
US-11-079-463-2940/c
; Sequence 2940, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 2940
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-2940

Query Match 12.1%; Score 108.8; DB 11; Length 1422;
Best Local Similarity 47.4%; Pred. No. 2.8e-12;
Matches 326; Conservative 0; Mismatches 362; Indels 0; Gaps 0;



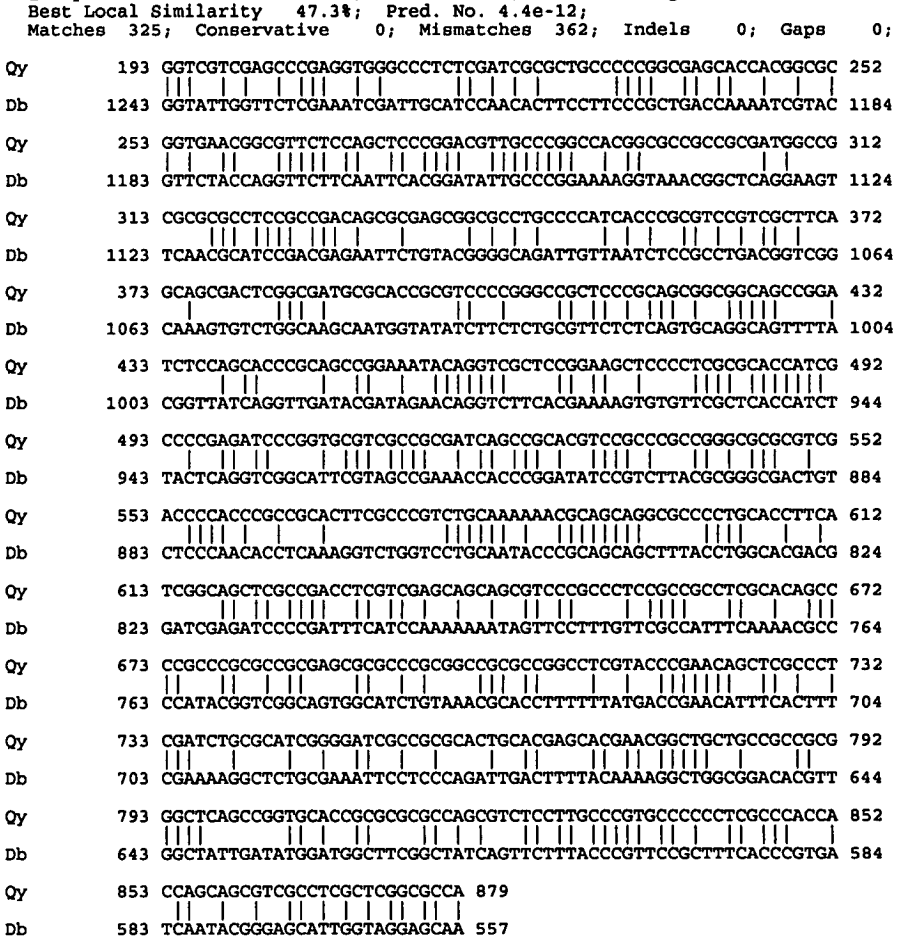
RESULT 11
US-11-079-463-3432/c
; Sequence 3432, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 3432
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-3432

Query Match 12.0%; Score 108; DB 11; Length 1017;
Best Local Similarity 46.9%; Pred. No. 4.3e-12;
Matches 336; Conservative 0; Mismatches 380; Indels 0; Gaps 0;



RESULT 12
US-11-079-463-549/c
; Sequence 549, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 549
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-549

Query Match 12.0%; Score 107.8; DB 11; Length 1434;
Best Local Similarity 47.3%; Pred. No. 4.4e-12;
Matches 325; Conservative 0; Mismatches 362; Indels 0; Gaps 0;



RESULT 13
US-11-079-463-291/c
; Sequence 291, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 291

```
;   LENGTH: 1395
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-291

  Query Match             11.5%;  Score 103.2;  DB 11;  Length 1395;
  Best Local Similarity   47.0%;  Pred. No. 3.3e-11;
  Matches  318;  Conservative    0;  Mismatches  358;  Indels    0;  Gaps    0;

Qy        221 GATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGAC 280
              ||| |    ||  ||    |  | || ||    | || ||  |  ||||  || ||||
Db       1203 GATGGTTTCACCATCGCTAATAATGACTGCTTTCTCAATGGAATGTTCCAATTCGCGGAT 1144

Qy        281 GTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGG 340
              |||||   ||| |      || |   |    | |||  ||  |||     |||   ||  |
Db       1143 ATTGCCATACCAGGCATGTGCGGTCAGTTTCTCGCAAGCGGCCGGACTCAGGCTGATAGA 1084

Qy        341 CGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTC 400
              ||  |   |  |             ||    |  |    ||| || |   | ||    ||
Db       1083 GGCTTTGTCATACTGTTTGCAAAAGCGGGCTATAAAACGCTCAGCTAGCGGAACAATATC 1024

Qy        401 CCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAG 460
              |   || |  || || || || |     |  |  |     |  ||| |  | |
Db       1023 TTCTTTACGTTTGCGAAGTGGAGGAATTTCTACATGAATGGTGTTGATACGGTATAATAA 964

Qy        461 GTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGAT 520
              ||   |||||    || | |   |||||| ||   |  | ||||  || ||   |||
Db        963 ATCTTCACGGAATAAGCCTTTGTCTACCATCTCCTGCAAGTTTCGGTTAGTTGCGCAGAT 904

Qy        521 CAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTG 580
              || || | |||  |  |     |  |  |    || || ||    |     | |  |||
Db        903 TAGGCGAATGTCTACGGGGATAGATTGGTTACTACCTACACGAACTATGCTTCTCTGCTG 844

Qy        581 CAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAG 640
              |      ||| ||     ||||||  |  |  || || |  || |  ||||| || |
Db        843 AATGGCTGTCAGTAGCTTAGCCTGCAAATGAAAGGGAAGGTTTCCTATTTCGTCAAGAAA 784

Qy        641 CAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGC 700
              ||| |  |   |  || || || ||   ||  || |  || | ||| | ||| | |
Db        783 CAGGGAACTTCGGTCTGCTGCTTCAAACTTACCTGTACGGTCGGCATGGGCGTCAGTAAA 724

Qy        701 CGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCA 760
              |  ||   | | |  |||||||| || | ||| | |   |  |||| ||| ||  |
Db        723 AGAACCTTTCACATGTCCGAACAGTTCACTCTCAAACAAGGATTCGGTGATAGCACCCAT 664

Qy        761 CTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAG 820
              |  |||   |  |   ||     || ||    |    || | | | || || ||
Db        663 ATCAACGGATATCATGCTCTCGGCAGAGCGTGGTGATAAAGCATGAATCTCACGTGCGAG 604

Qy        821 CGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCAC 880
              | | ||||| ||||| ||    || ||    | ||| |   |||| | | |||  || ||
Db        603 CATTTCCTTTCCCGTTCCGTTTTCACCTGTTATCAGTATATTCGCATTGGTCGTTGCTAC 544

Qy        881 CTTGCGCACCTGCGCG 896
              ||    | |  || |
Db        543 TTTCTCAATCAACGTG 528


RESULT 14
US-11-075-185-50
; Sequence 50, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 50
;   LENGTH: 765
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-50

  Query Match             11.2%;  Score 101.2;  DB 14;  Length 765;
  Best Local Similarity   62.0%;  Pred. No. 9.2e-11;
  Matches  160;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy        626 GACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGC 685
              || |||||||| | |||||| ||||| || ||| ||||| ||   ||   |||
Db          1 GATCTCGTCGAGGAACAGCGTGCCGCCGTCGGCCTCCTCGAACCTGCCGATCCGGAGGCG 60

Qy        686 GAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATC 745
              |||||  | |   ||||| ||||||   |||||||||||  ||||||   || | ||
Db         61 CGCGGCGCCGGTGAAGGCGCCCGCCTCGGCGCCGAACAGCTCCGCCTCGACGAGCTCGTC 120

Qy        746 GGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTG 805
              |||||  |||  ||   | |||   |||||||||| | | ||||||  |         ||
Db        121 GGGGAGCGCGCCGGCGTTCACCTTCACGAACGGCCGGTCCCGCCGGCGCGAGTTGGCCTG 180

Qy        806 CACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGC 865
              ||   ||  || ||  |||||||||| || ||    |   ||    | |||| || | |
Db        181 GACGATGCCGGCGAGGAGCTCCTTGCCGGTCCCGTTCGGCCCGGTGATGAGCACCGGCAC 240

Qy        866 CTCGCTCGGCGCCACCTT 883
              ||    ||||| |||||
Db        241 GTCCGCGGGCGCGACCTT 258


RESULT 15
US-11-075-185-2
; Sequence 2, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
;   LENGTH: 14172
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-2

  Query Match             11.2%;  Score 101.2;  DB 14;  Length 14172;
  Best Local Similarity   62.0%;  Pred. No. 4.6e-11;
  Matches  160;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy        626 GACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGC 685
              || |||||||| | |||||| ||||| || ||| ||||| ||   ||   |||
Db          1 GATCTCGTCGAGGAACAGCGTGCCGCCGTCGGCCTCCTCGAACCTGCCGATCCGGAGGCG 60

Qy        686 GAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATC 745
              |||||  | |   ||||| ||||||   |||||||||||  ||||||   || | ||
Db         61 CGCGGCGCCGGTGAAGGCGCCCGCCTCGGCGCCGAACAGCTCCGCCTCGACGAGCTCGTC 120

Qy        746 GGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTG 805
              |||||  |||  ||   | |||   |||||||||| | | ||||||  |         ||
Db        121 GGGGAGCGCGCCGGCGTTCACCTTCACGAACGGCCGGTCCCGCCGGCGCGAGTTGGCCTG 180
```

```
Qy        806 CACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGC 865
              ||   ||  || ||   ||||||||| || ||   |   ||    |  |||| || | |
Db        181 GACGATGCCGGCGAGGAGCTCCTTGCCGGTCCCGTTCGGCCCGGTGATGAGCACCGGCAC 240

Qy        866 CTCGCTCGGCGCCACCTT 883
              ||    ||||| |||||
Db        241 GTCCGCGGGCGCGACCTT 258
```

Search completed: April 27, 2006, 15:52:55
Job time : 566.152 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 4460.52 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_1_900
Perfect score: 900
Sequence: 1 tgggtatcccgagccgctgg.....cttgcgcacctgcgcgaaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	257	28.6	1135	9	BZ552870	BZ552870 pacs1-60_
	2	229.6	25.5	788	9	BZ578406	BZ578406 msh2_5838
c	3	223.2	24.8	1167	9	BZ578965	BZ578965 msh2_6086
	4	222.6	24.7	1081	9	BZ548914	BZ548914 pacs1-60_
c	5	216.2	24.0	841	7	CN822599	CN822599 Oa_splbn_
c	6	214.4	23.8	1079	9	BZ558461	BZ558461 pa98401_1
c	7	195.8	21.8	939	9	BZ569899	BZ569899 msh2_107.
c	8	187.8	20.9	1106	9	BZ558319	BZ558319 pacs1-60_
c	9	186.6	20.7	869	10	CL663408	CL663408 PRI0144a_
c	10	186	20.7	807	9	BZ553174	BZ553174 pacs1-60_
	11	182.6	20.3	1543	10	CL422970	CL422970 L001_PA_0
c	12	182.4	20.3	960	9	BZ573726	BZ573726 msh2_3333
	13	182.2	20.2	792	8	DN639706	DN639706 UMC-bemiv
	14	180.4	20.0	1237	9	BZ550946	BZ550946 pacs1-60_
	15	180.2	20.0	638	8	DN147566	DN147566 4917_F05_
c	16	171.2	19.0	1163	9	BZ548440	BZ548440 pacs1-60_
c	17	170.8	19.0	1049	9	BZ573814	BZ573814 msh2_3373
c	18	169.4	18.8	798	9	BZ566013	BZ566013 pacs2-164
	19	168.6	18.7	852	9	BZ571317	BZ571317 msh2_1830
c	20	164.8	18.3	1317	9	BZ572831	BZ572831 msh2_2820
c	21	163	18.1	1035	9	BZ561478	BZ561478 pacs2-164
c	22	161.8	18.0	743	9	BZ549766	BZ549766 pacs1-60_
	23	160.6	17.8	865	7	CN822849	CN822849 Oa_splbn_
	24	160.2	17.8	874	9	BZ573891	BZ573891 msh2_3412
c	25	156.6	17.4	647	9	BZ550651	BZ550651 pacs1-60_
	26	156.4	17.4	728	10	CL656058	CL656058 PRI0125c_
	27	152.4	16.9	1429	9	BZ575268	BZ575268 msh2_438.
c	28	151.2	16.8	1402	9	BZ576381	BZ576381 msh2_4908
c	29	149	16.6	1024	9	BZ563029	BZ563029 pacs2-164
c	30	146.8	16.3	1010	9	BZ551487	BZ551487 pacs1-60_
	31	145	16.1	820	7	CO202792	CO202792 Oa_splbn_
	32	144.4	16.0	632	8	DN147595	DN147595 4918_B05_
	33	144.2	16.0	707	10	CL682887	CL682887 PRI0135b_
	34	141.2	15.7	634	6	CB065384	CB065384 EST645065
	35	140.2	15.6	1140	9	BZ550221	BZ550221 pacs1-60_
c	36	139.8	15.5	820	9	AF075798	AF075798 AF075798
c	37	139.6	15.5	839	7	CN821727	CN821727 Oa_splbn_
c	38	139.6	15.5	869	7	CN823928	CN823928 Oa_splbn_
	39	138.8	15.4	1047	9	BZ573445	BZ573445 msh2_3202
c	40	137.2	15.2	789	9	BZ566021	BZ566021 pacs2-164
	41	137.2	15.2	1454	9	BZ569114	BZ569114 pacs2-164
	42	133.8	14.9	772	10	CL686556	CL686556 PRI0144c_
	43	133.4	14.8	508	8	DN641231	DN641231 UMC-bof_0
	44	133.2	14.8	680	10	CL683101	CL683101 PRI0135d_
c	45	133	14.8	1280	10	CL470317	CL470317 SAIL_140_

ALIGNMENTS

RESULT 1
BZ552870/c
LOCUS BZ552870 1135 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs1-60_3860.y2 pacs1-60 Pseudomonas aeruginosa genomic clone pacs1-60_3860, genomic survey sequence.
ACCESSION BZ552870
VERSION BZ552870.1 GI:27158079
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1135)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .1135
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_3860"
/clone_lib="pacs1-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."
ORIGIN

Query Match 28.6%; Score 257; DB 9; Length 1135;
Best Local Similarity 62.2%; Pred. No. 2.4e-35;
Matches 404; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

```
Qy        240 AGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCC 299
              ||  || | || || |  | |||||| ||||| || || |||| ||||| ||| ||
Db        750 AGATCCGCCGCCCGCTCGATGGCGTTTTCCAGTTCGCGCACGTGGCCCGCCCATGGGTAA 691
```

```
Qy        300 GCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGC 359
              ||| |||       || ||       |    ||||| |||  ||||||||
Db        690 TGGCGGATCGCCTGTTCGGCGTCGTGGACGAAGCTGAGCGACGCGCGCCCCATCTGGGTG 631

Qy        360 GTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGC 419
              | | ||   ||| | |    ||||||   || | ||||   |   ||||| |||||
Db        630 CACTGGCGAGCCAGAAAGGCGCGGGCGATCTCCATCACGTCGTTGCCACGCTCGCGCAGG 571

Qy        420 GGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCC 479
              | ||||   | |   | |||  |||| ||| |||| |||||     || ||    |||
Db        570 TCGGTCAGCTTCAGCGAGATCACATGCAGGCGGTAATAGAGGTCTTCGCGCAATTGCCCG 511

Qy        480 TCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGC 539
                 |||  |     || || ||||| ||  | ||||||| ||||||||   ||
Db        510 GTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTGTCGGCGATCAGGCGCACGTCGACCTTT 451

Qy        540 CGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGC 599
              | |   ||  ||| || ||||| || |  |||||   |||||| |  ||||||||||
Db        450 TGCGATTGCACCGAGCCGACCCGGCGGATCTCGCCTTCCTGCAATACGCGCAGCAGGCGG 391

Qy        600 CCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCC 659
              ||||  ||| || |||||| ||||||| ||||||||| | ||| || ||||| || |||
Db        390 GCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCGTCGAGGAACAGGGTACCGCCGTCGGCC 331

Qy        660 GCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCG 719
              || |||  |||||| || ||| | |||   || || | |   |||||    || | ||||
Db        330 GCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCACCGGTGAAGGCGCCTTTTTCATGCCCG 271

Qy        720 AACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGC 779
              |||||||||    |||||| | |   || || |||||||| || | |||     |  | ||||
Db        270 AACAGCTCGGATTCGATCAGGGTTTCCGGAATCGCCGCACAGTTCACCGAGATCAGCGGC 211

Qy        780 TGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCC 839
              ||    |||      |||  |||||  ||||| || | |   |||||||| |||||
Db        210 GCCTTGGCGCGCTTCGACAGGTTGTGCAGGGCGCGGGCGACCAGTTCCTTGCCGGTGCCG 151

Qy        840 CCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCA 888
              |||||||   | ||| | ||| |  ||| |||||||| |||||||| |
Db        150 GACTCGCCCTGGATCAGTACCGTGGAGTCGGTCGGCGCGACCTTGCGAA 102


RESULT 2
BZ578406
LOCUS       BZ578406                 788 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_5838.x1 msh Pseudomonas aeruginosa genomic clone msh2_5838,
            genomic survey sequence.
ACCESSION   BZ578406
VERSION     BZ578406.1  GI:27213467
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_5838"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
ORIGIN

  Query Match             25.5%;  Score 229.6;  DB 9;  Length 788;
  Best Local Similarity   60.2%;  Pred. No. 1.6e-30;
  Matches  399;  Conservative    0;  Mismatches  260;  Indels    4;  Gaps    1;

Qy        238 CGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCG 297
              || || || |  || |||  |||| |     ||||       ||    || ||  |
Db         28 CGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGC 87

Qy        298 CCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCC 357
              |||  |   ||  |||   |  | |   || | || |   |   |  | | |
Db         88 ACGCAACTTCTGCGCCGCCTCCGGGCTCAACTGGCTCTGGCCGTAGCCCAGGCGACGGCT 147

Qy        358 GCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCA 417
              |  | || || |  ||  |||  || |||    |||    ||| ||||| ||||| ||||
Db        148 GTATTCGGCGATGAAGTGGCGGGTCAGCG----GCAGGATGTCGCCGGGTCGCTCGCGCA 203

Qy        418 GCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCC 477
              ||||| |||   ||      | ||     |||||| | ||||||||     ||||||  ||
Db        204 ACGGCGACAGTTCGAGACTGACGACATTGAGCCGGTAGTACAGGTCCTCGCGGAAGTGCC 263

Qy        478 CCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCC 537
              |  |  | | |  | || | |  ||  |||||||  | ||  ||||||||
Db        264 CGGCATTGATCGCCCGCTCCAGTTGCACGTTGGTCGCCGCCAGCACGCGCACGTCGATGG 323

Qy        538 GCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGC 597
              |  |          |||||||| |||  |||| |  ||||||  |  |||||||
Db        324 GGATGCTCTTGCGCGACCCCAGGCGCACCACCTCGCGCTCCTGCAACACGCGCAGCAACT 383

Qy        598 GCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCG 657
              |||| | |  |||||||||| ||||||| |||||||| | |||| || ||||| | |
Db        384 TGACCTGGATCGGCATCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGTGCCGCCGTTGG 443

Qy        658 CCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACC 717
              || ||||| ||   || |||   ||||| || ||||| | |   || ||   |||||  |
Db        444 CCTCCTCGAACCAGCCGGCCTTGGCCGCCAGGGCGCCGGTGAAGGCACCCTTCTCGTGGC 503

Qy        718 CGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACG 777
              |||||||||| | || | | |  |||| ||  ||| |||| |  ||  ||||||||
Db        504 CGAACAGCTCGGCTTCCACCAGTGACTCGGAGAAGGCCCCGCAGTTGACCGCCACGAACG 563

Qy        778 GCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGC 837
              |    | ||||||| ||| ||   ||| |      ||||||| |   |||||||| ||||
Db        564 GGCCGTTGCGCCGCGCGCTGAGGTTGTGGATATGCCGCGCCACCAGTTCCTTGCCGGTGC 623

Qy        838 CCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGA 897
              |   |||||| | ||||| ||   | ||||||||||||||||| |||||  ||  |||
Db        624 CGGTCTCGCCGATCACCAACACGCTGGCCTCGCTCGGCGCCACCTGGCGCANGTGTTCGA 683

Qy        898 ACA 900
              |||
Db        684 GCA 686


RESULT 3
BZ578965/c
LOCUS       BZ578965                 1167 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_6086.x1 msh Pseudomonas aeruginosa genomic clone msh2_6086,
            genomic survey sequence.
ACCESSION   BZ578965
VERSION     BZ578965.1  GI:27214026
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1167)
```

```
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1167
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_6086"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
ORIGIN

  Query Match             24.8%;  Score 223.2;  DB 9;  Length 1167;
  Best Local Similarity   59.5%;  Pred. No. 2e-29;
  Matches  392;  Conservative    0;  Mismatches  266;  Indels    1;  Gaps    1;

Qy        242 CACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGC 301
              ||  ||  ||| |  ||  |||||||||| ||||| ||||  ||||||| ||| ||     |
Db        744 CANCAGCGCGTGTGGANTGGCGTTCTCAAGCTCGCGGATATTGCCCGCCCA-GGAATGCC 686

Qy        302 CGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGT 361
              ||  |   |||| || || ||    |||           |  |  |  |    ||||
Db        685 CGAGCAGCTTGCGCTCGGCTTCGGGCTCAGGCTCGCGCTGGTGTAGCNGAGGCGCCGCCA 626

Qy        362 CCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGG 421
              |     |  |  |||||||  |   |||      || ||| | |||| |||| |||
Db        625 GTGCTCGACGATGAAGTACTCGGCCAGCGGCAGGATATCGCCGCGGCGCTCGCGCAACGG 566

Qy        422 CGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTC 481
              ||||||  ||||   |  |||  |||| ||| |  |||||||    ||||||   ||  |
Db        565 CGGCAGTTGGATGGTCGCCACGTGCAGGCGGTAGAACAGGTCCTCGCGGAAGTGTCCGGC 506

Qy        482 GCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCG 541
              |  ||||    |  | || ||    ||  |||||||| | ||||||||||||     |
Db        505 GACCACCGCGTCAGCCAGGTCGACGTTGGTCGCCGCCACCAGCCGCACGTCGATGGGGAT 446

Qy        542 GGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCC 601
              |  ||   ||| || | ||||  ||| ||||     |||||  |  ||||||||     |
Db        445 CGGCCGGCGCGATCCGAGCCGCACCACCTCGCGTTCCTGCAGCACGCGCAGCAGCTTGAC 386

Qy        602 CTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGC 661
              |||  |   ||  |||||| |||||||  |||||||||  | |||  || |||||  |  || ||
Db        385 CTGGATGCTCGGCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGTGCCGCCGTTGGCTGC 326

Qy        662 CTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAA 721
              ||||  ||   ||  |||     |   | ||| || |     |||||   |||||  |||||
Db        325 CTCGAACCAGCCGGCCTTGCTCTGCAACGCACCGGTAAAGGCGCCCTTCTCGTGGCCGAA 266

Qy        722 CAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTG 781
              ||||||||  |||  | |    |||| ||  ||  |||| |  ||    ||||||||
Db        265 CAGCTCGCTCTCCACCAGGCTCTCGGCGAAAGCACCGCAGTTGACCGCGACGAACGGGCC 206

Qy        782 CTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCC 841
                  ||   |||  ||   |||  |    ||||||||  |  |||||||| |||||
Db        205 GCCGTTGCGTCCGCTGAGGGCGTGGATATGGCGCGCCACCAGCTCCTTGCCGGTGCCGGT 146

Qy        842 CTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||||||  |  ||||||    |  |||||||| |||||||          |  ||  ||||
Db        145 CTCGCCGATGACCAGCGCGTTGGCCTCGCTGGGCGCCAGGCGCTCGATGCGCTCCAACA 87
```

```
RESULT 4
BZ548914
LOCUS       BZ548914               1081 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_1559.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_1559, genomic survey sequence.
ACCESSION   BZ548914
VERSION     BZ548914.1  GI:27152495
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1081)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1081
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_1559"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             24.7%;  Score 222.6;  DB 9;  Length 1081;
  Best Local Similarity   58.3%;  Pred. No. 2.6e-29;
  Matches  390;  Conservative    0;  Mismatches  279;  Indels    0;  Gaps    0;

Qy        230 GCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGG 289
              |||  ||  |||  |  ||  ||||  |  ||        || |||||||| |||||||| ||
Db        126 GCCTTCGCAGAGGAGTACCGCGCATTCCACCAGTCCCTTCAGCTCCCGAACGTTGCCGGG 185

Qy        290 CCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCC 349
                  |     |  |||       |  |||| ||||| |||     | ||||||||  | |
Db        186 AAAGCCATAGCCGTCGAGCAGTTCCAGCGCTTCCGCGGACCAGCCCAGCGGCGCGCGGCG 245

Qy        350 CATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCG 409
              ||  | |           ||||      |     | ||| |      | ||||| ||    ||
Db        246 CAGGAACACGCAGGCGCGCTCGGCGAAATGCCGGGCCAGTTCGAGGGCGTCGCCCTCGCG 305

Qy        410 CTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCG 469
              ||  |||| |||||||| |  |||     | |  |  |||||| | || |||||     ||
Db        306 CTGGCGCAACGGCGGCAACTCGATGGGAAACTGCGCCAGCCGGTAGTAGAGGTCTTCGCG 365

Qy        470 GAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCAC 529
              |||  | || |||||  |||| |    |||| || ||||| || || ||||| |  || ||
Db        366 GAACTTGCCATCGCGGACCAGCAGGGCGAGGTCACGGTGGGTAGCGGCGATGATGCGTAC 425

Qy        530 GTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACG 589
              ||  ||    | | |     || | | | || |  |||||  |||||| |  ||
Db        426 ATCGACCTTGTGCGTGTCGTTGCCGCCGAGCGGGCGAATCTCGCCTTCCTGCAACACGCG 485

Qy        590 CAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCC 649
              || |||     |||| | |  ||  |||| |||||||  ||||||||||  ||| || ||
Db        486 CAACAGCTTGGCCTGGAGCGACAGGGGCATGTCGCCGATTTCGTCGAGCAACAGGGTTCC 545

Qy        650 GCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGC 709
              ||| ||||| || ||| ||| ||| |  ||  |       || || |     |||||
```

```
Db        546 GCCATCCGCGGCATCGAACAACCCGGGACGATCGCGCTCGGCACCGGTAAAGGCGCCCTT 605

Qy        710 CTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAG 769

Db        606 GCGGTAGCCGAACAGCTCGCTTTCCAGCAGGTTTTCCGGAAACGCCGCGCAGTTCTGCAA 665

Qy        770 CACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTT 829

Db        666 CACGAAATGCTGTGCACGCCGCGGGCCGTACTCGTGGATCGCCCGTGCGACCAGCTCCTT 725

Qy        830 GCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCAC 889

Db        726 TGCGGTTCCCGTCTCGCCCTGCAGCAACACGGTATAGGGGTTGTCCAAGACCTTGCTGAT 785

Qy        890 CTGCGCGAA 898

Db        786 CAGGGGGCA 794


RESULT 5
CN822599/c
LOCUS       CN822599                841 bp    mRNA    linear   EST 02-JUN-2004
DEFINITION  Oa_splbn_11B21_M13reverse Sheep spleen\brain pSport1 library Ovis
            aries cDNA clone Oa_splbn_11B21 5', mRNA sequence.
ACCESSION   CN822599
VERSION     CN822599.1  GI:47950668
KEYWORDS    EST.
SOURCE      Ovis aries (sheep)
  ORGANISM  Ovis aries
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Gossner,A. and Hopkins,J.
  TITLE     Ovine spleen\brain cDNA library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: J Hopkins
            Veterinary Biomedical Sciences
            University of Edinburgh
            Summerhall Square, Edinburgh, EH9 1QH.
            Email: j.hopkins@ed.ac.uk
            Plate: 11  row: B  column: 21
            Seq primer: M13reverse
            High quality sequence start: 4
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Ovis aries"
                     /mol_type="mRNA"
                     /db_xref="taxon:9940"
                     /clone="Oa_splbn_11B21"
                     /clone_lib="Sheep spleen\brain pSport1 library"
                     /note="Vector: pSport1"
ORIGIN

  Query Match             24.0%;  Score 216.2;  DB 7;  Length 841;
  Best Local Similarity   57.9%;  Pred. No. 3.5e-28;
  Matches  383;  Conservative    0;  Mismatches  278;  Indels    0;  Gaps    0;

Qy        240 AGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCC 299

Db        835 AGAATCAATGCACGCTGAATCGCATTGTCCAGTTCCCGCACGTTGCCGGGCCATGGGTAG 776

Qy        300 GCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGC 359

Db        775 CTGATCAGGCAGGCCTGAGCGTCGGCCGACAGGCGCGCCTGGGCATGCTTCATTTTATTG 716

Qy        360 GTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGC 419

Db        715 ACGTGCTTGCTCAGCAGGCGTTCGGCCAGCGGAATGATGTCTGCAGGGCGCTCGCGCAAC 656

Qy        420 GGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCC 479

Db        655 GGGCGCCAGGCCAACGGGAACACCGAGAGGCGGTAGAACAGATCCTCCCGGAAGCGCCCC 596

Qy        480 TCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGC 539

Db        595 GCCGCCACTTCGCCCGCCAGGTCGCGGTTGGTGGTGGCCACCACACGGATGTCCAGTTGG 536

Qy        540 CGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGC 599

Db        535 ATCGGCTTGCGTGCCCCCACCCGCTCCACTTCGCGCTCCTGCAATACCCGCAACAATTTG 476

Qy        600 CCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCC 659

Db        475 GCTTGCAGGCCAAGGGGCATTTCGGAGATCTCGTCCAGCAGGATGGTGCCGCCGTCGGCT 416

Qy        660 GCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCG 719

Db        415 TGTTCGAACTTGCCGGCCTGGGCGGCGATGGCGCCGGTGAACGAGCCTTTCTCATGGCCG 356

Qy        720 AACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGC 779

Db        355 AACAAGGTGGCTTCGAGCATGTTGTCGGGGATGGCCGCGCAGTTGATCGCAATGAACGGC 296

Qy        780 TGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCC 839

Db        295 TGACTGGCACGATGGGACTGCTGGTGGATGTAGCGCGCCAGCACCTCCTTGCCGGTGCCG 236

Qy        840 CCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAAC 899

Db        235 GACTCGCCGGAGATCAATACGGTGGAGTCGCTGCGGGCGACACGGGCCGCCAGCTCCAGC 176

Qy        900 A 900

Db        175 A 175


RESULT 6
BZ558461/c
LOCUS       BZ558461                1079 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pa98401_148.s1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pa98401_148, genomic survey sequence.
ACCESSION   BZ558461
VERSION     BZ558461.1  GI:27172874
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1079)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1079
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pa98401_148"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN
```

```
  Query Match             23.8%;  Score 214.4;  DB 9;  Length 1079;
  Best Local Similarity   63.5%;  Pred. No. 7.1e-28;
  Matches  343;  Conservative    0;  Mismatches  196;  Indels    1;  Gaps    1;

Qy        228 CTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCC 287
              |||| | ||  ||| | ||| || || |  | |||||| || || || || |||||||||
Db        587 CTGCTCTCGCTGAGGATCACCGCCCGCTCGATGGCGTTTTCGAGTTCGCGCACGTTGCCC 528

Qy        288 GGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGC 347
              ||||| ||         ||| |||      || ||    |  |    |||||||||   ||
Db        527 GGCCATGGGTAATGGCGGATCGCCTGTTCGGCGTCGTGGGCGAAGCTGAGCGGCGCGCGC 468

Qy        348 CCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCG-GG 406
              ||||||          | | ||   ||| |  |    ||||||   || | ||||    |
Db        467 CCCATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCCATCACGTCGTTGCCA 408

Qy        407 CCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCT 466
              |||||  |||||  |  ||||||    | |    | |||  |||| ||| |||| |||||
Db        407 GCGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGGTAATAGAGGTCTTC 348

Qy        467 CCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCG 526
              |||||   |||        |||  |     || || ||||| || || ||||||||  ||
Db        347 GCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTGGCAGCGATCAGGCG 288

Qy        527 CACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAA 586
              ||||||  ||  | | |   ||  ||| || ||||| || |  |||||   |||||| |
Db        287 CACGTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCGCCTTCCTGCAATAC 228

Qy        587 ACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGT 646
              ||||||||||  ||||  ||| || |||||| ||||||| ||||||||| | ||| ||
Db        227 GCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCGTCGAGGAACAGGGT 168

Qy        647 CCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCC 706
              ||||| || ||||| |||  |||||| || ||| | | |   || || | |   |||||
Db        167 GCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGAGCTGGCACCGGTGAAGGCGCC 108

Qy        707 GGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCAC 766
              || |  ||||||||||||   |||||| | |  || || |||||||| || | |||
Db        107 TTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTCCGGAATCGCCGCACAGTTCAC 48
```

```
RESULT 7
BZ569899/c
LOCUS       BZ569899                939 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_107.y1 msh Pseudomonas aeruginosa genomic clone msh2_107,
            genomic survey sequence.
ACCESSION   BZ569899
VERSION     BZ569899.1  GI:27204960
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 939)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .939
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_107"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
ORIGIN
```

```
  Query Match             21.8%;  Score 195.8;  DB 9;  Length 939;
  Best Local Similarity   63.0%;  Pred. No. 1.3e-24;
  Matches  302;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        230 GCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGG 289
              || | ||  ||| | ||| || || | || |||||| ||||| || || ||||||||||
Db        565 GCTCTCGTTGAGGATCACCGCCCGCTCAATGGCGTTTTCCAGTTCGCGCACGTTGCCCGG 506

Qy        290 CCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCC 349
              ||| ||    |  ||| |||      || ||    |  |    |||||||||  ||||
Db        505 CCATGGGTAATGCCGGATCGCCTGTTCGGCGTCGTGGGCGAAGCTGAGCGGCGCGCGCCC 446

Qy        350 CATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCG 409
              ||||         | | ||   ||| |  |    ||||||   || | ||||    |  ||
Db        445 CATCCGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCCATCACGTCGTTGCCACG 386

Qy        410 CTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCG 469
              ||| |||||  |  ||||||    | |    | |||  |||| ||| |||| |||||     ||
Db        385 CTCGCGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGGTAATAGAGGTCTTCGCG 326

Qy        470 GAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCAC 529
              |||   |||         ||  |     || |  ||||| || || |||||||| |||||
Db        325 GAATTGCCCGGTCTTGGACAGGGTTTTCAGGTGTCGGTGGGTGGCGGCGATCAGGCGCAC 266

Qy        530 GTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACG 589
              |||  ||  | | |   ||  ||| || ||||| || |  |||||   ||||||  |  ||
Db        265 GTCGACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCGCCTTCCTGCAATACGCG 206

Qy        590 CAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCC 649
              |||||||||  ||||  ||| || |||||| ||||||| |||||||| | ||| || ||
Db        205 CAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCGTCGAGGAACAGGGTGCC 146

Qy        650 GCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGG 708
              ||| || ||||| |||  |||||| || ||| | |||   || || | |   |||  ||
Db        145 GCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCACCGGTGAAGGCGGGGG 87
```

```
RESULT 8
BZ558319/c
LOCUS       BZ558319                1106 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_912.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_912, genomic survey sequence.
ACCESSION   BZ558319
VERSION     BZ558319.1  GI:27172488
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1106)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1106
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
```

```
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_912"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             20.9%;  Score 187.8;  DB 9;  Length 1106;
  Best Local Similarity   57.8%;  Pred. No. 3.3e-23;
  Matches  353;  Conservative    0;  Mismatches  257;  Indels    1;  Gaps    1;

Qy        168 GAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGAGGTGGGCCCTCTCGATCGCG 227
              | | | |   | |||   |    |||  ||  ||| |  |||| |||   || ||||  |
Db        641 GGGGCCGCGAGCGGCATGCCGATCGGCGCGGTGAGGCACAGGTCGGCGGGCTGGATCAGG 582

Qy        228 CTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCC 287
              | ||||      || | ||  || || || | ||||| || || ||||| |||||||||
Db        581 CCGCCCTGTTGCAGGATCAACGCACGCTGGATGGCGTTGTCGAGTTCCCGTACGTTGCCC 522

Qy        288 GGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGC 347
              ||||| |    |    | | |  |   |  ||||| |     |||||| || |   | |
Db        521 GGCCAGGCATGCCGGACCAGGCACTGGGCGGCCTCGGGGCCGAGCGCCACCGCGCCGAGG 462

Qy        348 CCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGC 407
                ||||   ||  |  || ||  ||||||   |||||| |   |||    |||| | ||
Db        461 TTCATCTTGCGGCTGTGTTGC-GCAGCAGACGCTCGGCCAGCGGCAGGATGTCCGCAGGG 403

Qy        408 CGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTC 467
              || || |||||||||| ||       | |   |  |||  ||| ||  | || |||||
Db        402 CGTTCGCGCAGCGGCCGCCAGGCCAGCGGGAAGACCGACAGGCGATAGTAGAGGTCCTCG 343

Qy        468 CGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGC 527
              ||||| | |||| |    ||   |||| |||| || ||||  ||||  || | ||  ||
Db        342 CGGAAACGCCCCGCCGCGACTTCCGCCGCGAGGTCGCGGTTGGTCGTGGCAAGCACGCGG 283

Qy        528 ACGTCCGCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAA 587
              | |||         |      ||  || ||||||  ||| |||| |  ||||||
Db        282 ATATCCAGGTTGATCGGTTTGCGCGCACCGACCCGCTCCACCTCGCGCTCCTGCAATACG 223

Qy        588 CGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTC 647
              |||| |||      ||||||      |  |||| ||     | |||||||| || |  ||
Db        222 CGCAACAGCTTGGCCTGCAGGCCGAGGGGCATTTCCGATATCTCGTCGAGAAGAATGGTG 163

Qy        648 CCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCCGCGCGAGCGCGCCCGCGGCCGCGCCG 707
              ||||| || |||  |||| ||   || | | | || || |  ||||| | |   |  ||
Db        162 CCGCCGTCGGCCAGCTCGAACTTGCCGGGCTGGGCGGCAATGGCGCCGGTGAAGGAACCT 103

Qy        708 GCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACG 767
              |||||  ||||||||   | ||||||  |       || ||||||||||||| | |
Db        102 TTCTCGTGGCCGAACAGGGTGGCCTCGAGCATGTTGTCCGGGATCGCCGCGCAGTTGATC 43

Qy        768 AGCACGAACGG 778
                | ||| ||
Db         42 GCGATGAAGGG 32


RESULT 9
CL663408/c
LOCUS       CL663408                869 bp    DNA     linear   GSS 09-JUL-2004
DEFINITION  PRI0144a_E11 - PRI0144a.B21 (869) Mixed stage fosmid library of P.
            pacificus var. California Pristionchus pacificus genomic, genomic
            survey sequence.
ACCESSION   CL663408
VERSION     CL663408.1  GI:50152447
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
  TITLE     AppaDB: an AcedB database for the nematode satellite organism
            Pristionchus pacificus
  JOURNAL   Nucleic Acids Res. 32 (1), D421-D422 (2004)
  PUBMED    14681447
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            This library was generated at Caltech, Pasadena, USA and end
            sequenced at Vancouver, Canada.
            Seq primer: T7
            Class: fosmid ends.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Mixed stage fosmid library of P. pacificus
                     var. California"
                     /note="Vector: pEpifos-5 Fosmid vector"
ORIGIN

  Query Match             20.7%;  Score 186.6;  DB 10;  Length 869;
  Best Local Similarity   55.7%;  Pred. No. 5.4e-23;
  Matches  357;  Conservative    0;  Mismatches  284;  Indels    0;  Gaps    0;

Qy        240 AGCACCACGGCGCGGTGAACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCC 299
              |||||||| || || |  || ||||| |||||||| || |  || ||||||||
Db        817 AGCACCACTGCCCGTTCCACCGCGTTTTCCAGCTCACGAATATTTCCCGGCCAGTCGTAA 758

Qy        300 GCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGC 359
                  |    |   |||||  | ||  |  |     |||||  |   |   |
Db        757 TGAATCAACAGATCCATCGCCTGGGGCGTAAAACCTTTTACCGCCTTGCGATTACGCTCG 698

Qy        360 GTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGC 419
              |    |||| ||| |       | || |   |         ||  |  |||| | |||||||
Db        697 GCAAAGCGCTGCAGAAAATGGCCAGCCAGCAGAGGGATATCTTCCCGCCGTTGCCGCAGC 638

Qy        420 GGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCC 479
              |  || |     ||| ||| |||   ||| ||  | || || ||    || || | |||
Db        637 GATGGTACTTCAATCGCCACCACATTCAGGCGATAGTAGAGATCCTGGCGAAAACGCCCG 578

Qy        480 TCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGC 539
              |   ||||   ||  | ||||| || || |||||||| |||||||| || ||  ||
Db        577 GCATTCACCTCTGCGGCAAGATCGCGATGGGTCGCCGCAATCAGCCGGACATCAACCGAG 518

Qy        540 CGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGC 599
              |   |  |     || || ||| | |||||||  | ||| |    |||||||||| |||
Db        517 ATAGTCTGGTTGCTACCGACACGCTGAACTTCGCGCTCCTGAATCGCACGCAGCAGACGC 458

Qy        600 CCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCC 659
              |||||| | ||  ||  |  ||||| |  || ||||| | |||||| ||||| |||||
Db        457 ACCTGCATCATCGGCGAGATATCGCCAATTTCATCGAGAAACAGCGTGCCGCCGTCCGCC 398

Qy        660 GCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCG 719
              |  |  |  ||||  ||||        || || |    ||| ||   ||||  |||
Db        397 TCAACAAAGCGCCCTTCCCGGCGTTTATCGGCTCCAGTAAACGCCCCTTTTTCGTGACCG 338

Qy        720 AACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGC 779
              ||||  ||    || | |   | |||  ||  ||||| || |  |     || | |||
Db        337 AACAATTCAGATTCCAGCAAGGATTCGTTGAGTGCCGCACAGTTGAGCGTTACCAGCGGT 278

Qy        780 TGCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCC 839
              |  |  |  || | |||       ||| | ||| |  || | |  ||| || || |||||
Db        277 TTTTCGCTACGTGCGCTACTGGCGTGAATCGCCCTGGCGACCAGCTCTTTACCGGTGCCG 218

Qy        840 CCCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCAC 880
              |||||    | ||||| ||| || ||    ||||| ||
Db        217 GAATCGCCGTGGATCAGCACCGTGGCTTCCGATGGCGCGAC 177
```

```
RESULT 10
BZ553174/c
LOCUS       BZ553174                807 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_4015.x1 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_4015, genomic survey sequence.
ACCESSION   BZ553174
VERSION     BZ553174.1  GI:27158871
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 807)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .807
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_4015"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
ORIGIN

  Query Match           20.7%;  Score 186;  DB 9;  Length 807;
  Best Local Similarity   65.0%;  Pred. No. 7e-23;
  Matches  335;  Conservative    0;  Mismatches  176;  Indels    4;  Gaps    4;

Qy        390 CGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCA-GCACCCGCAG 448
              | || | ||||      ||||| ||||| |  ||||||   | |    |  |||  ||||
Db        727 CCCATCACGTCGGTTCCACGCTCGCGCAGGGCGGGCAGCTTCAGCGAGATCCACATGCAG 668

Qy        449 CCGGAAATACAGGTC-GCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGT 507
              ||| |||| |||||  | ||||||   |||        |||  |     || || ||||
Db        667 NCGGTAATAGAGGTCTTCGCCGGAATTGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGT 608

Qy        508 GCGTCGCCGCGATCA-GCCGCACGTCCG-CCCGCCGGGCGCGCGTCGACCCCACCCGCCG 565
              | || || ||||||| | |||||||| | ||  | | |   ||  ||| || ||||| ||
Db        607 GGGTGGCGGCGATCAGGGCGCACGTCGGACCTTCTGCGATTGCACCGAGCCGACCCGGCG 548

Qy        566 CACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCC 625
              | |||||   |||||| |  ||||||||||  ||||  ||| || |||||| |||||
Db        547 GATCTCGCCTTCCTGCAATACGCGCAGCAGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCC 488

Qy        626 GACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGC 685
              || |||||||||| | ||| || ||||| || ||||| |||  |||||| || ||| | ||
Db        487 GATCTCGTCGAGGAACAGGGTGCCGCCGTCGGCCGCTTCGACCAGCCCGGCGCGCCCGGC 428

Qy        686 GAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATC 745
              |    || || | |   |||||     || |  ||||||||||||    |||||| | |  ||
Db        427 GCTGGCACCGGTGAAGGCGCCTTTTTCATGGCCGAACAGCTCGGATTCGATCAGGGTTTC 368

Qy        746 GGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCGCGGGCTCAGCCGGTG 805
              || |||||||| || | |||     | | ||||  ||     |||   |||    |||
Db        367 CGGAATCGCCGCACAGTTCACCGAGATCAGCGGCGCCTTGGCGCGCTTCGACAGGTTGTG 308

Qy        806 CACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCACCACCAGCAGCGTCGC 865
              ||  ||||| || | |.. |||||||| |||||   |||||||   | ||| | ||| |
```

```
Db        307 CAGGGCGCGGGCGACCAGTTCCTTGCCGGTGCCGGACTCGCCCTGGATCAGTACCGTGGA 248

Qy        866 CTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
              ||| ||||||| |||||||| | ||    | |||
Db        247 GTCGGTCGGCGCGACCTTGCGAATCTTGCTGTACA 213


RESULT 11
CL422970
LOCUS       CL422970                1543 bp    DNA     linear   GSS 09-APR-2004
DEFINITION  L001_PA_0004_A03-0004_L07 pooled genomic library of 12 strains of
            Pseudomonas aeruginosa Pseudomonas aeruginosa genomic clone
            L001_PA_0004_A03-0004_L07, genomic survey sequence.
ACCESSION   CL422970
VERSION     CL422970.1  GI:45388101
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1543)
  AUTHORS   Shen,K., Sayeed,S., Hu,F.Z., Antalis,P., Gladitz,J., Ahmed,A.,
            Hayes,J., Johnson,S., Dopico,R., Preston,R.A., Erdos,G., Post,J.C.
            and Ehrlich,G.D.
  TITLE     Sequence from pooled genomic library of 12 isolates of Pseudomonas
            aeruginosa
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Shen K
            Center for Genomic Sciences
            Allegheny-Singer Research Institute
            320 East North Avenue, Pittsburgh, PA 15212, USA
            Tel: 412 359 5097
            Fax: 412 359 6995
            Email: kshen@wpahs.org
            Complete insert sequence
            Seq primer: M13 Reverse, and T7
            Class: shotgun
            High quality sequence stop: 1543.
FEATURES             Location/Qualifiers
     source          1. .1543
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="pooled"
                     /db_xref="GDB:1111034"
                     /db_xref="taxon:287"
                     /clone="L001_PA_0004_A03-0004_L07"
                     /cell_type="bacteria"
                     /clone_lib="pooled genomic library of 12 strains of
                     Pseudomonas aeruginosa"
                     /note="Vector: pCR4-TOPO; using TOPO Shotgun method; 12
                     strains include: 1) 11 isolates from clinical samples of
                     pediatric  otitis media with effusion: Pitt 'A', Pitt 'B',
                     Pitt 'C',Pitt 'D',  Pitt 'E', Pitt 'F', M18851, M18858,
                     W27912, W27931, and W28869;  2) 1 strain from American
                     Type Culture Collection (ATCC): 27853 "
ORIGIN

  Query Match           20.3%;  Score 182.6;  DB 10;  Length 1543;
  Best Local Similarity   59.2%;  Pred. No. 2.6e-22;
  Matches  311;  Conservative    0;  Mismatches  214;  Indels    0;  Gaps    0;

Qy        371 CAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCG 430
              |||||||  ||||  |    || |  ||| ||    |||||  |  ||||||||||    |
Db         35 CAGCAGCGCTTCGGACAGCAACAGCAGGTCACCTTCGCGCTCTCTGAGCGGCGGCAAGTG 94

Qy        431 GATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCAT 490
                |   |   |||     |  ||  |  ||| |||    |||||||  |||||  |   ||
Db         95 CAGTACGCCCACAGCGATGCGATAGAACAAGTCTTCACGGAAGCGCCCCTGGGCGACTTC 154

Qy        491 CGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGT 550
              |  |    || |  ||||| || || | ||||||  |||||  ||  ||     |  | | |
Db        155 CTGCATCAGGTTGCGGTGGGTGGCGGTGATCAGTCGCACATCGACCTTGAGCTCCTGATT 214
```

```
Qy        551 CGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTT 610
Db        215 TGAACCGACCGGAGTGAAGGTGCCTTCCTGCAGTACACGCAGCAGTCGCACCTGTACATC 274

Qy        611 CATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAG 670
Db        275 TGGCGTCAGCTCACCGAACTCATCGAGAAACAGCGTGCCGCCGTGAGCCTGCTGGAACAC 334

Qy        671 CCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCC 730
Db        335 CCCAGGCCGATCGGCAATCGCGCCAGTGAATGCGCCCTTCTTGTGGCCAAACAGGGTGGA 394

Qy        731 CTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTGCTGCCGCCG 790
Db        395 GTCGATCAGCTCCGGCACGATGGCACCGCAGTTCACCGCAACAAAGGGCGCTGCAGAACG 454

Qy        791 CGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCCCCTCGCCCAC 850
Db        455 CGGGGCGGCGTTGTGAATGGCGCGGGCAAACAACTCTTTGCCGGTGCCGGTTTCACCGTA 514

Qy        851 CACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGC 895
Db        515 GATCAGCACCGGCACCTGCTTCGCCGCGAGTATTTGCGCCTGCGC 559
```

RESULT 12
BZ573726/c
LOCUS BZ573726 960 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_3333.y2 msh Pseudomonas aeruginosa genomic clone msh2_3333, genomic survey sequence.
ACCESSION BZ573726
VERSION BZ573726.1 GI:27208787
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 960)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .960
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_3333"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun library."
ORIGIN

Query Match 20.3%; Score 182.4; DB 9; Length 960;
Best Local Similarity 61.2%; Pred. No. 2.9e-22;
Matches 294; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

```
Qy        421 GCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCT 480
Db        565 GGGGCGGTTCCAACTTCAGGGCGTTTAACCGGTAGTATAAAACCTCTCGGAAAGTTCCGG 506

Qy        481 CGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCC 540
Db        505 CATCGATCGCCGCCTTCAGGTTGACATGGGTTACCGGCAGCAACCGGACGTCCACCTCGA 446

Qy        541 GGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCC 600
Db        445 ACGGCCGGCTACCGCCGAACCGCTCGATTTGCTGTTCCTGGAGAAAGCGCAACAGGTTGG 386

Qy        601 CCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCG 660
Db        385 CCTGCAACTCCAGCGGCAGGTCGCCGATCTCGTCGAGGAACAGGGTGCCGCCGTCGGCCG 326

Qy        661 CCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGA 720
Db        325 CCTCGATCCGCCCGACCTTGCGCTGGTGCGCGCCGGTGAAGGCGCCCTTCTCGTGGCCGA 266

Qy        721 ACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCT 780
Db        265 ACAGCTCGGACTGGATCAGGTGCTCGGGAATCGCCCCGCAGTTGATCGCGACGAACGGCC 206

Qy        781 GCTGCCGCCGCGGGCTCAGCCGGTGCACCGCGCGCGCCAGCGTCTCCTTGCCCGTGCCCC 840
Db        205 TGTCCGCGCGCCGCGACTGCCGGTGCAGGGTACGCGCGACCAGTTCCTTGCCGGTGCCGC 146

Qy        841 CCTCGCCCACCACCAGCAGCGTCGCCTCGCTCGGCGCCACCTTGCGCACCTGCGCGAACA 900
Db        145 TCTCGCCCCGGATCAGTACCGGCGAGTCGGTCGGCGCCAGCTTCGCCAGCAGCCGGCGCA 86
```

RESULT 13
DN639706
LOCUS DN639706 792 bp mRNA linear EST 28-MAR-2005
DEFINITION UMC-bemiv_0A01-026-e06 Early Embryo In Vitro Produced bemiv Bos taurus cDNA 3', mRNA sequence.
ACCESSION DN639706
VERSION DN639706.1 GI:61944366
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 792)
AUTHORS Prather,R.S., Antoniou,E., Garverick,H.A., Green,J.A., Lucy,M.C., Roberts,R.M., Smith,M.F. and Youngquist,R.S.
TITLE USDA Grant NRI-2002-03476: Bovine ESTs: Focus on Female Reproduction
JOURNAL Unpublished (2002)
COMMENT Contact: DNA Core Facility (Bovine Project)
Animal Science - RS Prather
University of Missouri-Columbia
M616 Medical Sciences Bldg., Columbia, MO 65212, USA
Tel: (573)882-0428
Fax: (573)884-5552
Email: bovine@rnet.missouri.edu
POLYA=No.
FEATURES Location/Qualifiers
source 1. .792
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone_lib="bemiv"
/note="Funding: The production of ESTs submitted in this project was funded by USDA Grant MRI-2002-03476 entitled 'Bovine ESTs: Focus on Female Reproduction' to RS Prather, E Antoniou, HA Garverick, JA Green, MC Lucy, RM Roberts, MF Smith and RS Youngquist. Genetic Source: Heifers for the project were purchased from Circle A Ranch, Iberia, MO (http://www.circlearanch.com/home.html). These heifers, while not registered have known Angus pedigrees going back at least 4 generations. Samples collected: The samples consisted of the following: germinal vesicle-stage oocytes; in vitro derived embryos (2-cell, morula, blastocyst and nuclear transfer blastocyst); in vivo blastocysts and conceptuses (days 8, 14, 16 and 18);

corpora lutea (days 3, 5, 8, 14, 16, 18 and 35); ovarian follicles (days 0, non-recruited, recruited, early selected and preovulatory); oviduct (days 0, 3 and 5); endometrium (days 5, 8, 14, 16, 18 and 35); and placenta/embryo from day 35 conceptuses. Expanded descriptions of how the tissues were collected can be found at the following URL: http://genome.rnet.missouri.edu/Bovine/Methods.html. Library construction (Standard Protocol): All procedures have been described in detail elsewhere (Soares et al., 1994; Bonaldo et al., 1996; Jiang et al., 2001). Total cellular RNA from each sample was isolated by using STAT-60 reagent (Tel-Test, Friendswood, TX) and the poly(A)+ RNA was obtained by two rounds of purification with the Oligotex mRNA isolation kit (Qiagen) according to the manufacturer''s instructions. The libraries were constructed by E. Ferguson and R. Woods essentially as described by the manufacturer's instructions provided with the SuperScript Plasmid System (Invitrogen, cat. no. 18248-013). Briefly, 1mcg of poly(A)+ RNA was annealed at c37 degrees with 10mcg of NotI-tag-dT18 oligonucleotide (GCTGCTCGCGGCCGC-tag-T18) and reverse transcribed at c37 degrees with SuperScript II (Invitrogen) reverse transcriptase (Jiang et al., 2001). The 'tag' represents a tissue/stage-specific ten-base sequence identifier (http://genome.uiowa.edu/pubsoft/software.html) present in the oligonucleotide used to prime first-strand synthesis. Second strand synthesis was performed with T4 DNA polymerase in the presence of DNA ligase and RNase H. After second strand synthesis, the double-stranded cDNAs were ligated to SalI adapters (Invitrogen-Life Technologies) and digested with NotI. The cDNAs were size selected by passage through cDNA size fractionation columns (Invitrogen-Life technologies). The cDNAs derived from each developmental stage of a particular tissue were mixed on an equimolar basis and ligated directionally into the NotI and SalI sites of the pCMV-SPORT6 vector (Invitrogen). After ligation of the inserts, the plasmids were electroporated into DH10B bacteria. Library Construction (PCR Protocol): The amount of mRNA that was recovered from oocytes and embryos was quite limiting and was not sufficient for library production with the standard protocol. Therefore, PCR-based protocol was utilized for producing libraries from sources in which the amount of extracted mRNA was small (oocytes and embryos). Poly-A RNA was isolated by using the MicroPoly(A) Pure kit from Ambion (cat. # 1918). The mRNA was reverse transcribed with a NotI-tag-dT18 oligonucleotide and a SMART oligonucleotide (Clontech) modified to contain a SalI site to generate full-length cDNA with a sequence complementary to the SMART oligonucleotide. Sequences within the SMART and dT oligonucleotides were used as primers to amplify the cDNAs by PCR with pfu turbo polymerase (Stratagene). The resulting PCR products were purified, digested with NotI and SalI and size fractionated by using Chroma Spin-1000 columns (Clontech). Purified cDNA from each PCR reaction was quantitated and mixed on an equimolar basis for ligation into the pCMV-SPORT6 vector. Preliminary Library Characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (~4 96-well plates) to confirm library quality [e.g. the presence of short polyA+ tails, genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.] and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. After production of the libraries, equal numbers of recombinants from each library were pooled to produce a single mixed library (mega-library) for more extensive sequencing. Bioinformatics work was performed by GK Springer's bioinformatics group (Spollen WG, Topinka CM, Khambati AA) in Computer Science at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core facility at: bovine@rnet.missouri.edu. Bonaldo MF, Lennon G, Soares MB,. Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Res, 1996; 6:791-806. Jiang H, Bivens NJ, Ries JE, Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion BA, Mathialagan N, Prather RS, Lucy MC (2001) Constructing cDNA libraries with fewer clones that contain long poly(dA) tails. Biotechniques 31:38-42. Soares MB, MF Banaldo, P Jelene, L Su, L Lawton, A Efstrantiadis. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci, 91:9228-9232. TAG_TISSUE=Early Embryo In Vitro Produced TAG_SEQ=Not found"

ORIGIN

Query Match 20.2%; Score 182.2; DB 8; Length 792;
Best Local Similarity 56.1%; Pred. No. 3.2e-22;
Matches 343; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

```
Qy      148 CGTCGCGCAGCGATGCGGAGGAGACGGGCGGAGGCGGCCGCGGCGGGTCGTCGAGCCCGA 207
            ||  | || ||| ||| |    | |||| | ||    | |||||       ||| ||
Db      180 CGCTGAGCGGCGGTGCCGCAAGGGTGGGCCGCGGTTCGCCAGGCGGACGCGGCAGCGGGA 239

Qy      208 GGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTGAACGGCGTTCT 267
             || | |    | | ||   |||  ||||||| |  || || || |  || || | |
Db      240 AGTCGTCGAGACCCAGCGTATCGCCGTCGGCGAGGATGACCGCTCGTTCGACCGCATGCC 299

Qy      268 CCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCG 327
            || | | ||||| ||||||||||| |      |    | |||  || ||| || | ||
Db      300 GCAACGCACGGACATTGCCCGGCCAGGAATAGTCGAGCAGCGCCCTGCTCGCATCGGGCG 359

Qy      328 ACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGA 387
             ||| |  |    | | ||       ||||           ||       |||  |
Db      360 TGAGCCCACGTTCGGGCCGTCCGTATTTGCGCGCATAGAAGCCGAGGTAGTGGTCGAGCA 419

Qy      388 TGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCA 447
             ||      ||| || | ||  |||||||||| ||||||  |||||||| |     ||
Db      420 GCAGCGGGATATCCTCGCGGCGGGCCCGCAGCGGGGGCAGCTCGATCTCCACCGTGTTCA 479

Qy      448 GCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCCGAGATCCCGGT 507
            | |||||   ||||||   |||||||| |.||||     || |  | ||        ||
Db      480 GTCGGAACAGCAGGTCCTGCCGGAAGCGGCTCTCGTCGTGCAGCCGCTCGGTGGTGGCGT 539

Qy      508 GCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGGCGCGCGTCGACCCCACCCGCCGCA 567
             || || ||||| |  |||||||   | ||  ||| ||  |     || ||| ||  ||
Db      540 TGGTGGCGGCGATGACGCGCACGTCGATCGGCACGGCCCGGTTGCTGCCGACCGGCGTCA 599

Qy      568 CTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGA 627
            |||  | |   |  |      ||||||  |||  |||||  | || ||||||  | |||||
Db      600 CTTGCCGCTGTTCGAGCGCGGTCAGCAGTCGCGGCTGCAGGTGCAGCGGCAGGTTGCCGA 659

Qy      628 CCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGA 687
            |||||||||  | ||| || | ||| || |||||||  | |  ||||    ||  |   |
Db      660 TCTCGTCGAGGAACAGGGTACTGCCGTCGGCCGCCTGGAAGCGCCCGATGCGTTCGCTGC 719

Qy      688 GCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGATCTGCGCATCGG 747
            |||||  | |   |||||     |||  |||||  ||||||   ||||  |  |||  ||
Db      720 GCGCGTCGGTGAAGGCGCCCTTGACGTGACCGAAGAGCTCGGAATCGAACAGCGATTCCT 779

Qy      748 GGATCGCCGCG 758
            |||  ||  ||
Db      780 GGACGGCGCCG 790
```

RESULT 14
BZ550946
LOCUS BZ550946 1237 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs1-60_2908.y2 pacs1-60 Pseudomonas aeruginosa genomic clone pacs1-60_2908, genomic survey sequence.
ACCESSION BZ550946

```
VERSION     BZ550946.1  GI:27154527
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1237)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_2908"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             20.0%;  Score 180.4;  DB 9;  Length 1237;
  Best Local Similarity   61.5%;  Pred. No. 6.4e-22;
  Matches  289;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy        294 GGCGCCGCCGCGATGGCCGCGCGCGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATC 353
              ||  || ||   || |||      || ||    |  |    ||||| |||  ||||||||
Db         91 GGATCCCCCTTTATCGCCTGTTCGGCGTCGTGGGCGAAGCTGAGCGACGCGCGCCCCATC 150

Qy        354 ACCCGCGTCCGTCGCTTCAGCAGCGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCC 413
                  | | ||   ||| | |    ||||||   || | ||||   |   |||||
Db        151 CGGGTGCACTGGCGAGCCAGAAAGGCGCGGGCGATCTCCATCACGTCGTTGCCACGCTCG 210

Qy        414 CGCAGCGGCGGCAGCCGGATCTCCAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAG 473
              ||||| |  ||||||   | |   | |||  |||| ||| |||| |||||    |||||
Db        211 CGCAGGGCGGGCAGCTTCAGCGAGATCACATGCAGGCGGTAATAGAGGTCTTCGCGGAAT 270

Qy        474 CTCCCCTCGCGCACCATCGCCCCGAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCC 533
                |||       |||  |     || || ||||| || || |||||||| ||||||||
Db        271 TGCCCGGTCTTGGCCAGGGTTTTCAGGTCGCGGTGGGTGGCGGCGATCAGGCGCACGTCG 330

Qy        534 GCCCGCCGGGCGCGCGTCGACCCCACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGC 593
                ||  | | |   ||  ||| || ||||| || |  |||||   ||||||  ||||||
Db        331 ACCTTCTGCGATTGCACCGAGCCGACCCGGCGGATCTCGCCTTCCTGCAATACGCGCAGC 390

Qy        594 AGGCGCCCCTGCACCTTCATCGGCAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCC 653
              |||||  ||||  ||| || |||||| |||||| ||||||||  | ||| || ||| |
Db        391 AGGCGGGCCTGGGCCTCCAGCGGCAGTTCGCCGATCTCGTCGAGGAACAGGGTGCCGGCG 450

Qy        654 TCCGCCGCCTCGCACAGCCCCGCCCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCG 713
              || ||||| |||  |||||| || ||| | |||   || || | |   |||||    ||
Db        451 TCGGCCGCTTCGACCAGCCCGGCGCGCCCGGCGCTGGCACCGGTGAAGGCGCCTTTTTCA 510

Qy        714 TACCCGAACAGCTCGCCCTCGATCTGCGCATCGGGGATCGCCGCGCACTG 763
              |  ||||||||||   |||||| | |   |    |||   |  || || |
Db        511 TGGCCGAACAGCTCGGATTCGATCAGGGTTTTTCGGAAATCGCCGGACAG 560


RESULT 15
DN147566
LOCUS       DN147566                638 bp    mRNA    linear   EST 16-FEB-2005
DEFINITION  4917_F05_K09 Switchgrass leaf cDNA library Panicum virgatum cDNA
            clone 4917_F05_K09, mRNA sequence.
ACCESSION   DN147566
VERSION     DN147566.1  GI:59868417
KEYWORDS    EST.
SOURCE      Panicum virgatum (switchgrass)
  ORGANISM  Panicum virgatum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Paniceae; Panicum.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   Tobias,C.M., Twigg,P., Hayden,D.M., Fladseth,M.R., Vivian,L.A.,
            Chow,E.K. and Sarath,G.
  TITLE     An EST survey of Switchgrass: a C4 perennial grass
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tobias CM
            Genomics and Gene Discovery Unit
            USDA, Agricultural Research Service, Western Regional Research
            Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 510 559-6172
            Fax: 510 559-5818
            Email: ctobias@pw.usda.gov
            The phred basecalling program was used to call bases and identify
            the high scoring region using the -trim_alt '' and trim_cutoff 0.01
            options. Vector sequences have been removed using the program
            cross_match.
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1. .638
                     /organism="Panicum virgatum"
                     /mol_type="mRNA"
                     /cultivar="Kanlow"
                     /db_xref="taxon:38727"
                     /clone="4917_F05_K09"
                     /dev_stage="immature tillers"
                     /lab_host="E. coli DH5alpha"
                     /clone_lib="Switchgrass leaf cDNA library"
                     /note="Organ: Leaf; Vector: pSPORT1; Site_1: Sal1; Site_2:
                     Not1; Switchgrass cv. Kanlow plants were grown from seed
                     in a greenhouse in potting soil. Immature tillers were
                     collected randomly, Leaf tissue was flash frozen in liquid
                     nitrogen. The tissue was prepared by Gautam Sarath, USDA,
                     ARS, Wheat Forage and Sorghum Research Unit, U. Nebraska,
                     Lincoln. Total RNA and poly(A) RNA were prepared, cDNA
                     synthesized, and directionally ligated into pSPORT1 by
                     Paul Twigg, Biology Department, U. Nebraska Kearney,
                     Kearney, NE. Plasmid DNA preparations and DNA sequencing
                     were performed in the laboratory of CM Tobias."
ORIGIN

  Query Match             20.0%;  Score 180.2;  DB 8;  Length 638;
  Best Local Similarity   56.8%;  Pred. No. 7.4e-22;
  Matches  332;  Conservative    0;  Mismatches  253;  Indels    0;  Gaps    0;

Qy        197 GTCGAGCCCGAGGTGGGCCCTCTCGATCGCGCTGCCCCCGGCGAGCACCACGGCGCGGTG 256
              | | || | || ||       || ||||  || ||||      |  | ||  || || ||
Db         33 GGCCAGGCAGAAGTCCTGGGGCTGGATCAAGCCGCCCTGCTGCAAAATCAACGCACGCTG 92

Qy        257 AACGGCGTTCTCCAGCTCCCGGACGTTGCCCGGCCACGGCGCCGCCGCGATGGCCGCGCG 316
              |  || || ||||| ||||| ||||||||| ||||| ||          | |   ||  |
Db         93 GATCGCATTGTCCAGTTCCCGCACGTTGCCGGGCCATGGGTAGCTGATCAGGCAGGCCTG 152

Qy        317 CGCCTCCGCCGACAGCGCGAGCGGCGCCTGCCCCATCACCCGCGTCCGTCGCTTCAGCAG 376
              || || ||||||||      | | || |||  |||           |      |||||||
Db        153 GGCGTCGGCCGACAGGCGCGCCTGGGCATGCTTCATTTTATTGACGTGCTTGCTCAGCAG 212

Qy        377 CGACTCGGCGATGCGCACCGCGTCCCCGGGCCGCTCCCGCAGCGGCGGCAGCCGGATCTC 436
                |||||| |   | |    |||  | || |||||  |||| |||   ||     | |
Db        213 GCGCTCGGCCAGCGGAATGATGTCTGCCGGGCGCTCGCGCAACGGGCGCCAGGCCAACGG 272

Qy        437 CAGCACCCGCAGCCGGAAATACAGGTCGCTCCGGAAGCTCCCCTCGCGCACCATCGCCCC 496
              | ||||   || ||| |  |||  ||    |||||||  |||| |   |||     || |
Db        273 GAACACCGAGAGGCGGTAGAACAAATCCTCACGGAAGCGCCCCGCCGCCACTTCGCCCGC 332
```

```
Qy        497 GAGATCCCGGTGCGTCGCCGCGATCAGCCGCACGTCCGCCCGCCGGGCGCGCGTCGACCC 556
              || || ||||  || |  || | ||  || | ||||    |    |        |  ||
Db        333 CAGGTCGCGGTTGGTGGTGGCAACCACACGGATGTCCAGTTGGATCGGCTTGCGTGCGCC 392

Qy        557 CACCCGCCGCACTTCGCCCGTCTGCAAAAAACGCAGCAGGCGCCCCTGCACCTTCATCGG 616
              |||||||  |||||||| |  |||||| |  |||||||      | ||||     |  ||
Db        393 CACCCGCTCCACTTCGCGCTCCTGCAATACCCGCAGCAACTTGGCTTGCAGGCCAAGGGG 452

Qy        617 CAGCTCGCCGACCTCGTCGAGCAGCAGCGTCCCGCCCTCCGCCGCCTCGCACAGCCCCGC 676
              ||  |||  ||  |||||||||||| |  || ||||| || |||   ||| ||   || ||
Db        453 CATTTCGGAGATTTCGTCGAGCAGGATGGTGCCGCCATCGGCCTGTTCGAACTTGCCGGC 512

Qy        677 CCGCGCCGCGAGCGCGCCCGCGGCCGCGCCGGCCTCGTACCCGAACAGCTCGCCCTCGAT 736
              | | || ||||  ||||| | |  || |||    || |  |||||||    | | ||||
Db        513 CTGGGCGGCGATGGCGCCGGTGAACGAGCCTTTTTCATGGCCGAACAAGGTGGCTTCGAG 572

Qy        737 CTGCGCATCGGGGATCGCCGCGCACTGCACGAGCACGAACGGCTG 781
              |      |||||||| |||||||| |  |     | |||||||||
Db        573 CATGTTGTCGGGGATGGCCGCGCAGTTGATCGCAATGAACGGCTG 617
```

Search completed: April 28, 2006, 12:17:39
Job time : 4463.52 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 5546.62 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_2000_3000
Perfect score: 1001
Sequence: 1 caggcgatgccggtgagcgg.....cccccgcgcaggagcgcgag 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1001	100.0	84868	1	DQ013294	DQ013294 Polyangiu
c 2	242.2	24.2	2973	6	AR628165	AR628165 Sequence
3	242.2	24.2	9867	6	AR619613	AR619613 Sequence
c 4	176	17.6	4957	1	AY929300	AY929300 Pseudomon
c 5	174.2	17.4	4377	6	AR148449	AR148449 Sequence
c 6	174.2	17.4	4377	6	AR177931	AR177931 Sequence
c 7	174.2	17.4	4377	6	CS131783	CS131783 Sequence
c 8	168.8	16.9	6979	1	AB045630	AB045630 Pseudomon
c 9	168.8	16.9	10929	1	AE004886	AE004886 Pseudomon
c 10	166.8	16.7	300933	1	AE016791	AE016791 Pseudomon
c 11	165.6	16.5	110000	1	CP000076_60	Continuation (61 o
12	160.4	16.0	110000	1	CP000075_09	Continuation (10 o
13	155.6	15.5	110000	1	AE016853_10	Continuation (11 o
14	138	13.8	110000	1	CP000058_10	Continuation (11 o
c 15	133.4	13.3	110000	1	CP000011_08	Continuation (9 of
16	132.4	13.2	1279	10	PM2H12G	AL684840 Penicilli
17	130.2	13.0	110000	1	BX571966_19	Continuation (20 o
c 18	129.8	13.0	1393	10	PM11H12G	AL684264 Penicilli
19	128.8	12.9	1393	10	PM11H12G	AL684264 Penicilli
20	127.6	12.7	65351	14	AC139773	AC139773 Homo sapi
c 21	125.4	12.5	33451	1	AY541063	AY541063 Nocardia
c 22	125.4	12.5	207592	14	CR392332	CR392332 Danio rer
23	123.8	12.4	159997	14	CR352334	CR352334 Danio rer
24	122.8	12.3	1094	10	PM7G11B	AL685196 Penicilli
c 25	122.6	12.2	1094	10	PM7G11B	AL685196 Penicilli
c 26	121	12.1	186752	8	AC111200	AC111200 Homo sapi
c 27	120.4	12.0	138390	13	AY261359	AY261359 Bovine he
28	119.2	11.9	1143	10	PM12A11G	AL684286 Penicilli
c 29	119	11.9	991	10	PM12H12B	AL684455 Penicilli
30	119	11.9	110000	1	BX571965_26	Continuation (27 o
31	119	11.9	110000	1	CP000010_16	Continuation (17 o
c 32	119	11.9	159997	14	CR352334	CR352334 Danio rer
33	116.6	11.6	136753	14	AC149968	AC149968 Strongylo
34	116.2	11.6	150715	13	AY714813	AY714813 Cercopith
c 35	116.2	11.6	150715	13	AY714813	AY714813 Cercopith
c 36	115.6	11.5	1279	10	PM2H12G	AL684840 Penicilli
c 37	115.2	11.5	110000	1	BX571965_05	Continuation (6 of
c 38	115.2	11.5	110000	1	CP000010_32	Continuation (33 o
c 39	114.4	11.4	10444	13	PVI422133	AJ422133 Suid herp
c 40	114.2	11.4	186620	14	CR405691	CR405691 Danio rer
c 41	113.6	11.3	136753	14	AC149968	AC149968 Strongylo
c 42	113.4	11.3	65351	14	AC139773	AC139773 Homo sapi
43	112.8	11.3	142014	14	CR407553	CR407553 Danio rer
44	112.6	11.2	1377	10	PM12C6G	AL684346 Penicilli
45	111.8	11.2	956	10	PM2D12B	AL684743 Penicilli

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRAREAEREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPPFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
CDS complement(1992. .3440)
/note="ORFu2; similar to Pseudomonas putida KT2440 PP4696; putative nitrogen regulation protein NR(I)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="AAY32957.1"
/db_xref="GI:63099934"
/translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
ELTETEIARRLGISRKALWERRQRLGIPRPRS"
CDS 3587. .5548
/note="ORFu3; ORFu3p; similar to Leptospira interrogans AAN50474"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAY32958.1"
/db_xref="GI:63099935"
/translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
VVAGFGATGPLRVMETLPEQYLFLTRVRP"
CDS complement(5702. .7231)
/note="ORFu4; ORFu4p; similar to Burkholderia fungorum hypothetical protein in ZP_00033665"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAY32959.1"
/db_xref="GI:63099936"
/translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPPSANQRPASTLAQWRVGLEHP
GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
GRVITVHPHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
P"
CDS complement(7561. .8931)
/note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440 PP4696; putative nitrogen regulation protein NR(I)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="AAY32960.1"
/db_xref="GI:63099937"
/translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
PPPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
IPRPRS"
CDS complement(8928. .11939)
/note="ORF1; ORF1p; similar to Pseudomonas putida KT2440 HisK (PP4695) in GenBank Accession Number AAN70268"
/codon_start=1
/transl_table=11
/product="putative sensory box histidine kinase"
/protein_id="AAY32961.1"
/db_xref="GI:63099938"
/translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
FTAFFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPAREEREAAR
LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
CDS complement(12405. .13379)
/note="ORF2; ORF2p; similar to family of known or putative phosphotransferases, including macrolide 2-phosphotransferases; similar to Bacillus subtilis YcbJ, Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273 MphB and Staphylococcus aureus plasmid pSR1 MphBM"
/codon_start=1
/transl_table=11
/product="putative macrolide kinase"
/protein_id="AAY32962.1"
/db_xref="GI:63099939"
/translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
CDS 13368. .14663
/note="ORF3; ORF3p"
/codon_start=1
/transl_table=11
/product="putative Ser/Thr protein kinase"
/protein_id="AAY32963.1"
/db_xref="GI:63099940"
/translation="MFGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
gene 15731. .33766
/gene="dszA"
CDS 15731. .33766
/gene="dszA"
/note="PKS (polyketide synthase) subunit modules 1-4a"
/codon_start=1
/transl_table=11
/product="DszA"
/protein_id="AAY32964.1"
/db_xref="GI:63099941"
/translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
PRAHYSPDPNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT

```
                     LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                     LSPDGLCKTFDADANGYVPGDGVGLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                     ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                     RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                     VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                     RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                     DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                     RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                     RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                     ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                     RAQFTFKKPLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                     LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                     PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                     PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                     EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE

  Query Match             100.0%;  Score 1001;  DB 1;  Length 84868;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-56;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9574 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 9633

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9634 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 9693

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9694 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 9753

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9754 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 9813

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9814 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 9873

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9874 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 9933

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9934 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 9993

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9994 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 10053

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10054 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 10113

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10114 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 10173

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10174 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 10233

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10234 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 10293

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10294 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 10353

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10354 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 10413

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10414 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 10473

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10474 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCC 10533

Qy        961 GCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001
              |||||||||||||||||||||||||||||||||||||||||
Db      10534 GCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 10574


RESULT 2
AR628165/c
LOCUS       AR628165               2973 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 9566 from patent US 6833447.
ACCESSION   AR628165
VERSION     AR628165.1  GI:59736977
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2973)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 9566 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .2973
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             24.2%;  Score 242.2;  DB 6;  Length 2973;
  Best Local Similarity   57.0%;  Pred. No. 1.4e-07;
  Matches  561;  Conservative    0;  Mismatches  388;  Indels   36;  Gaps    5;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              | |||||||  | ||  ||||||||||||  | |||  ||||||||||||||||||  |  |||||
Db       2330 CTGGCGATGGCCGTCAGCGGGTTGCCAATCTCATGGGCCACGCCCGCGGCCACGCGCCCC 2271

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              |    |||||| |||||  | ||  ||  |    ||||  ||  |   ||      |  |||
Db       2270 AGTGAGGCCAGCCGGTCCTGATGCGCCATGCGCGCGTCCACGGCCTTTCGTGCCGTGAAG 2211

Qy        121 TCCTCCACGAGCAGGACCACGC---CGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGC 177
              |||||||  ||  |||  |||  ||   ||  ||||  |  |  |  |   ||    |||  ||||
Db       2210 TCCTCCACCAGGAGGGCCATGCCCTCGGACGAGCCCGTGCCCTGCTCCGCGGGAGAGAGC 2151

Qy        178 GCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCG 237
                |  |||||||  ||  ||  |  ||||  |    ||  ||||||   ||     |||||  |
Db       2150 CGCGAGCGGTGCAGCCGGAGAATGCGCTCGACGCCCGCCACGGTGACACGTGTCTCCGTG 2091

Qy        238 GCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCG 297
              | |                ||    |  |||  |||    |  ||  |  ||||    |
Db       2090 TCTC-----------------CGTCCACACCGCTGGCGAAACCGGACAGCAGCGGGCCC 2049

Qy        298 AACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACG 357
              | |||||  |||||     ||  ||  ||     ||     |  |  |  |  ||    |||||  |
Db       2048 CAGGGGGCGGGCAGGGCATCCAGGGGATGGCCACGCACGGTGTCCACCGGCACGCCGGAG 1989

Qy        358 AGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACC 417
              |||  ||||||    |  |||   ||  |  ||  |  |  ||  ||  |  ||||||||||||||
Db       1988 AGCTGCTCGAGCGTCGCGTTCCAGATGACGACCTCACCATCCGGCCCCACGGCGCAGACG 1929

Qy        418 CCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTC 477
              ||  |  ||||  |  |  ||  |     ||||||  |  ||  ||||  |    ||  |||  |   |
Db       1928 CCCACCGGGAGGTCCTCCAGGATGCGGCGCAGGTAGCGGCGCACGGCCTCCAGCGCGGGC 1869
```

```
Qy        478 GCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCC 537
Db       1868 ACCGGGCCTTGCAGTCCCCGCGCGTCGCG---AAGGCGCTCCTCGACGAAGCGGAGTTGT 1812

Qy        538 TCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCC 597
Db       1811 TCGGCCAGGGCGGTGCGCGCGC---------CGGGCTCCAGCCGCAGCGCCTCCTCCACC 1761

Qy        598 GCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGC 657
Db       1760 GTTATGCGCGCGAGCACGGGGCCCAGCAGTCCCGACAGGTTTCGCTCCACGCCGTCGCGC 1701

Qy        658 AGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717
Db       1700 AGGCGGCGCAGCTCGGAGGGCCGGCGCTCGTCCTGCGGCAGCGACAGGGAATTCAGCGCC 1641

Qy        718 CGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGC 777
Db       1640 CGGTCCACCTCCGCCGTCGCGGCCTCCTCACCCAGCAGCGGGGCCAGCTTGCGGCGGAAC 1581

Qy        778 GCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTG 837
Db       1580 TCG---TCGGGCGAGCCCGCCACCACGCCGCCCGCGGCGGGCGCGGCCTCCTCGCGTGTG 1524

Qy        838 CACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCG 897
Db       1523 CAGGCCCGCGCGGCCTCTGCTTCCTCGGCGGACTGGCGCGTCGCCAGCGACACGCCGATG 1464

Qy        898 AGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGC 957
Db       1463 AAGGCCAGCCCGTTGAGCGTGAGCGAGGCGAAGGTGGCGAAGCCCCAGGGCTCACTGTCG 1404

Qy        958 GCCGCGCCCGCCGGCGCGCCCGCGC 982
Db       1403 GAGAACCCGAGCAGCGCCGCCATGC 1379
```

```
RESULT 3
AR619613
LOCUS       AR619613                9867 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 1008 from patent US 6833447.
ACCESSION   AR619613
VERSION     AR619613.1  GI:59724019
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 9867)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 1008 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .9867
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

```
  Query Match           24.2%;  Score 242.2;  DB 6;  Length 9867;
  Best Local Similarity   57.0%;  Pred. No. 8e-08;
  Matches  561;  Conservative    0;  Mismatches  388;  Indels   36;  Gaps    5;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
Db       1543 CTGGCGATGGCCGTCAGCGGGTTGCCAATCTCATGGGCCACGCCCGCGGCCACGCGCCCC 1602

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
Db       1603 AGTGAGGCCAGCCGGTCCTGATGCGCCATGCGCGCGTCCACGGCCTTTCGTGCCGTGAAG 1662

Qy        121 TCCTCCACGAGCAGGACCACGC---CGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGC 177
Db       1663 TCCTCCACCAGGAGGGCCATGCCCTCGGACGAGCCCGTGCCCTGCTCCGCGGGAGAGAGC 1722
```

```
Qy        178 GCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCG 237
Db       1723 CGCGAGCGGTGCAGCCGGAGAATGCGCTCGACGCCCGCCACGGTGACACGTGTCTCCGTG 1782

Qy        238 GCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCG 297
Db       1783 TCTC------------------CGTCCACACCGCTGGCGAAACCGGACAGCAGCGGGCCC 1824

Qy        298 AACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACG 357
Db       1825 CAGGGGGCGGGCAGGGCATCCAGGGGATGGCCACGCACGGTGTCCACCGGCACGCCGGAG 1884

Qy        358 AGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACC 417
Db       1885 AGCTGCTCGAGCGTCGCGTTCCAGATGACGACCTCACCATCCGGCCCCACGGCGCAGACG 1944

Qy        418 CCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTC 477
Db       1945 CCCACCGGGAGGTCCTCCAGGATGCGGCGCAGGTAGCGGCGCACGGCCTCCAGCGCGGGC 2004

Qy        478 GCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCC 537
Db       2005 ACCGGGCCTTGCAGTCCCCGCGCGTCGCG---AAGGCGCTCCTCGACGAAGCGGAGTTGT 2061

Qy        538 TCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCC 597
Db       2062 TCGGCCAGGGCGGTGCGCGCGC---------CGGGCTCCAGCCGCAGCGCCTCCTCCACC 2112

Qy        598 GCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGC 657
Db       2113 GTTATGCGCGCGAGCACGGGGCCCAGCAGTCCCGACAGGTTTCGCTCCACGCCGTCGCGC 2172

Qy        658 AGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717
Db       2173 AGGCGGCGCAGCTCGGAGGGCCGGCGCTCGTCCTGCGGCAGCGACAGGGAATTCAGCGCC 2232

Qy        718 CGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGC 777
Db       2233 CGGTCCACCTCCGCCGTCGCGGCCTCCTCACCCAGCAGCGGGGCCAGCTTGCGGCGGAAC 2292

Qy        778 GCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTG 837
Db       2293 TCG---TCGGGCGAGCCCGCCACCACGCCGCCCGCGGCGGGCGCGGCCTCCTCGCGTGTG 2349

Qy        838 CACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCG 897
Db       2350 CAGGCCCGCGCGGCCTCTGCTTCCTCGGCGGACTGGCGCGTCGCCAGCGACACGCCGATG 2409

Qy        898 AGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGC 957
Db       2410 AAGGCCAGCCCGTTGAGCGTGAGCGAGGCGAAGGTGGCGAAGCCCCAGGGCTCACTGTCG 2469

Qy        958 GCCGCGCCCGCCGGCGCGCCCGCGC 982
Db       2470 GAGAACCCGAGCAGCGCCGCCATGC 2494
```

```
RESULT 4
AY929300/c
LOCUS       AY929300                4957 bp    DNA     linear   BCT 12-JUL-2005
DEFINITION  Pseudomonas putida MhaS (mhaS) and MhaR (mhaR) genes, complete cds.
ACCESSION   AY929300
VERSION     AY929300.1  GI:62860911
KEYWORDS    .
SOURCE      Pseudomonas putida
  ORGANISM  Pseudomonas putida
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 4957)
  AUTHORS   Arias-Barrau,E., Sandoval,A., Naharro,G., Olivera,E.R. and
            Luengo,J.M.
  TITLE     A two-component hydroxylase involved in the assimilation of
```

```
          3-hydroxyphenyl acetate in Pseudomonas putida
  JOURNAL   J. Biol. Chem. 280 (28), 26435-26447 (2005)
   PUBMED   15866873
REFERENCE   2  (bases 1 to 4957)
  AUTHORS   Arias-Barrau,E., Luengo,J.M. and Olivera,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2005) Bioquimica Y Biologia Molecular,
            Universidad De Leon, Campus De Vegazana S/N, Leon 24007, Spain
FEATURES             Location/Qualifiers
     source          1. .4957
                     /organism="Pseudomonas putida"
                     /mol_type="genomic DNA"
                     /strain="U"
                     /db_xref="taxon:303"
     gene            421. .3396
                     /gene="mhaS"
     CDS             421. .3396
                     /gene="mhaS"
                     /note="two-component signal sensor"
                     /codon_start=1
                     /transl_table=11
                     /product="MhaS"
                     /protein_id="AAY16575.1"
                     /db_xref="GI:62860912"
                     /translation="MPMSFSLTQMILISAGYLMVLFGVAWISERGLIPRSIIRHPLTY
                     TLSLGVYASAWAFYGSVGLAYQYGYGFLACYLGVSGAFLLAPVLLYPILKITRTYQLS
                     SLADLLAFRFRSTWAGALTTIIMLIGVLPLLALQIQAVADSISILTGEPVKARVAFAF
                     CTLIILFTIFFGSRHIATREKHEGLVFAIAFESVIKLLALGGIGLYALYGVFGGPHEL
                     EVWLLQNQTALAALHTPLQEGPWRTLLLVFFASAIVMPHMYHMAFTENLNPRSLVSAS
                     WGLPLFLLLMSLAVPLVLWAGLRLGATTNPEYFTLGLGIAANNEALALLAYVGGLSAA
                     SGLIIVTTLALSGMALNHLVLPLYQPPAEGNIYRWLKWTRRALIVAIITAGFMFYLTQ
                     NNHQSLANLGIVAFVATLQFLPGVLSVLYWPTANRRGFIAGPLAGTLVWMVTMLLPLL
                     GNLQGFYIPLLDMIYVLDGTSWRMAAIASLAANVLLFTLISLFTNASTEEVSAAEACA
                     VDNVRRPQRRELHAASPQEFATQLAKPLGAKAAQKEVEQALRDLYLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDMVETFLPYKSGNENYVTEDIHFIESRLEDYHSRLTGLA
                     AELDALRRYHRQTLQELPMGVCSLAKDQEILMWNKAMEELTGIAAKHVVGSRLLTIDE
                     PWRGLLQGFINLPDEHLHKQRLALDGQPRWLNLHKPAIDEPLAPGNSGLVLLVEDLTE
                     TQALEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGEIIELSSQ
                     ILDQTKRVSRIVQSLMSFAHAGGSHQNSEEPVCLAEVAQDAIGLLALNRRNFEVQFFN
                     LCNPDHWAEGDPQRLAQVLINLLSNARDASPPGSAVRVRSEVSEHTVDLIVEDEGSGI
                     PKNIMDRLFEPFFTTKDPGEGTGLGLALVYSIVEEHYGQITIDSPADIERQRGTRIRV
                     TLPRHVVATSPEIRDRREN"
     gene            3401. .4843
                     /gene="mhaR"
     CDS             3401. .4843
                     /gene="mhaR"
                     /note="two-component response regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="MhaR"
                     /protein_id="AAY16576.1"
                     /db_xref="GI:62860913"
                     /translation="MPHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERFSIA
                     TFDLIVSDLRLPGAPGTELIKLGQGTPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILRDRQNAPAAASAAEPRATNGKAASADKGSAAAANGEIGIIGSCPPMQ
                     DMYSKIRKVAPTDSNVLIQGESGTGKELVARALHNLSRRAKAPMISVNCAAIPETLIE
                     SELFGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRV
                     GSVQSQKVDVRLIAATHRDLKNLAKAGQFREDLYYRLHVIALKLPALRERGSDVNEIA
                     SAFLARQSARIGRDDLHFSAEAEQAIRHYSWPGNVRELENAVERAVILSESAEISAEL
                     LGIDIELSDLEEDEILNNALVVASAASASHEPTEDLSLEDYFQHFVLEHQDHMTETEL
                     ARKLGVSRKCLWERRQRLGIPRRKSNATSE"
ORIGIN

  Query Match           17.6%;  Score 176;  DB 1;  Length 4957;
  Best Local Similarity  53.2%;  Pred. No. 0.0017;
  Matches  502;  Conservative    0;  Mismatches  415;  Indels   27;  Gaps    5;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||| | |||||||||||  |||||||||||||| ||||| || || |||
Db       2750 CAGGCGATGCCGGTAATCGGGTTGCCGATTTCGTGGGCCACGCCGGCGGCCAGGCGGCCG 2691

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
                ||   ||||||||| |||  |||  | ||||   | || || ||  |    |||||  |

Db       2690 ATGCTGGCCAGGCGCTCGGAGTGCACCAGCTTGTCTTCCAGGGCCTGTGTTTCGGTAAGG 2631

Qy         121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
               || ||||| ||||| || | |||||      |||   | ||      ||   ||||  |||
Db       2630 TCTTCCACCAGCAGCACAAGGCCGCTGTTACCGG---GGGCCAACGGCTCGTCGATGGCG 2574

Qy         181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
               | |  ||||| |  |||   ||||| | ||||| |||           |  |    ||  |
Db       2573 GGCTTGTGCAGGTTCAGCCAGCGCGGCTGCCCGTCCA----------GTGCCAGGCGCTG 2524

Qy         241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
               |  | || ||  |    || | |||  |      ||  ||     | |||  ||    ||
Db       2523 CTTGTGCAGGTGTTCGTCGGGCAGGTTGATGAAACCTTGC-----AGCAGGCCCCGCCAC 2469

Qy         301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
               ||  |   |   |  ||  |||  ||||||| ||      |  ||  ||||   |||
Db       2468 GGTTCGTCGATGGTGAGCAGGCGCGAGCCGACGACGTGCTTGGCGGCGATGCCGGTCAGC 2409

Qy         361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
               | || | |  | |||     |    |||  |   |||   | |||  | ||||||  ||
Db       2408 TCTTCCATGGCCTTGTTCCACATCAGGATTTCCTGGTCCTTGGCCAGCGAGCAGACGCCC 2349

Qy         421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
               |  || |||| |   ||   | |||  |  | || ||||||  |||||||||     |||
Db       2348 ATGGGCAGCTCCTGCAGGGTCTGGCGGTGGTAACGGCGCAGGGCATCGAGCTCGGCGGCC 2289

Qy         481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
                 ||||   || ||||  |    |   | |||||  |  || |   |  |||  || |||
Db       2288 AGGCCGGTCAGGCGCGAGTGATAGTCTTCCAGGCGGCTTTCGATGAAGTGGATGTCTTCG 2229

Qy         541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
               ||   | ||   |    ||   ||||||  |  ||||  |   |  | ||  ||   |
Db       2228 GT---CACGTAGTTTTCGTTACCGGACTTGTACGGCAGGAAGGTTTCGACCATGTC---- 2176

Qy         601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
                  ||| |   ||| || | |||     |||||||||   |  |  || ||  ||||
Db       2175 --CTGCGCCACGCTCGGCCCCATCAGGCCTGACAGGTTGGCCTCGATGCGGTCGCGCAGG 2118

Qy         661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
                 | |   ||  ||| ||   |||||||  ||  | || ||| ||   | ||||  |
Db       2117 CGACGCAAGGCGTAGGGGCGACGCTCGTCGAACGGCAGGTAGAGGTCACGCAGGGCCTGT 2058

Qy         721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
               || ||  |   ||| |||| |  ||| ||| ||  || ||  |    | | | |
Db       2057 TCCACTTCCTTTTGGGCGGCCTTGGCGCCCAGCGGCTTGGCCAGTTGGGTAGCGAACTCC 1998

Qy         781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
                  |   |  ||    | |||||  |  || | |||  | | |    | |||
Db       1997 TGCGGCGAGGCGGCGTGCAGCTCGCGCCGCTGCGGGCGGCGTACGTTGTCCACGGCGCAG 1938

Qy         841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
               ||||||   |  |||||  ||  |    |  | |||||||| |   || |
Db       1937 GCTTCGGCCGCGCTGACTTCCTCTGTGCTGGCGTTGGTGAACAGCGAGATCAGGGTGAAC 1878

Qy         901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
                  |   ||||| | || ||||  | | | ||  |    |||
Db       1877 AGCAAAACGTTGGCCGCCAGCGAGGCAATGGCCGCCATGCGCCA 1834


RESULT 5
AR148449/c
LOCUS       AR148449               4377 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 28 from patent US 6225106.
ACCESSION   AR148449
VERSION     AR148449.1  GI:15112539
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4377)
  AUTHORS   Gerritse,G. and Quax,W.J.
  TITLE     Expression system for altered expression levels
  JOURNAL   Patent: US 6225106-A 28 01-MAY-2001;
```

```
FEATURES             Location/Qualifiers
     source          1. .4377
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             17.4%;  Score 174.2;  DB 6;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 0.0024;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              |||||  || |||||  ||||  |||||     |  |  ||  ||||     |    ||||  ||
Db        632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  |||||  |   ||   ||||  ||   |||  |||    |||   |    ||   |    |
Db        575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              |  |  ||  |||  |  |   |  ||  ||                  | ||||||      |
Db        515 TGCTCCTCGTCCTGGGCGATGAAGTCC--------------TGCAGCAGCTCGCGCCAG 471

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||  |  |||  |  |   ||  |||  |||||  |   |   ||  |   |  ||||  |||
Db        470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||  |  |||||     |       |    |||    | |||   |  ||  |  ||||  |||
Db        410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  ||  ||  |  |   ||   |  |||   |  |  ||  |||||     ||||||||      |||
Db        350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  |||  ||||  ||  |      |  |  |||||  |  |||  |    |||   ||  |||
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              |||     | |       ||||  |     |  |  |  |  ||  ||     |||   |||
Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              ||  |  |  ||  |  |||||          ||  ||  |||    |   |  ||  ||   ||||
Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              |  |  ||  ||  |||  ||   |||||||    |  |  ||  ||  |||   |  ||||  |||
Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
              ||  ||  ||   ||  |  |  |   |||  ||  ||   ||  |||   |  |  |  |
Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 6
AR177931/c
LOCUS       AR177931               4377 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 28 from patent US 6313283.
ACCESSION   AR177931
VERSION     AR177931.1  GI:17920286
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.

REFERENCE   1  (bases 1 to 4377)
  AUTHORS   Gerritse,G. and Quax,W.J.
  TITLE     Expression system for altered expression levels
  JOURNAL   Patent: US 6313283-A 28 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .4377
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             17.4%;  Score 174.2;  DB 6;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 0.0024;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              |||||  || |||||  ||||  |||||     |  |  ||  ||||     |    ||||  ||
Db        632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  |||||  |   ||   ||||  ||   |||  |||    |||   |    ||   |    |
Db        575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              |  |  ||  |||  |  |   |  ||  ||                  | ||||||      |
Db        515 TGCTCCTCGTCCTGGGCGATGAAGTCC--------------TGCAGCAGCTCGCGCCAG 471

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||  |  |||  |  |   ||  |||  |||||  |   |   ||  |   |  ||||  |||
Db        470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||  |  |||||     |       |    |||    | |||   |  ||  |  ||||  |||
Db        410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  ||  ||  |  |   ||   |  |||   |  |  ||  |||||     ||||||||      |||
Db        350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  |||  ||||  ||  |      |  |  |||||  |  |||  |    |||   ||  |||
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              |||     | |       ||||  |     |  |  |  |  ||  ||     |||   |||
Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              ||  |  |  ||  |  |||||          ||  ||  |||    |   |  ||  ||   ||||
Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              |  |  ||  ||  |||  ||   |||||||    |  |  ||  ||  |||   |  ||||  |||
Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
              ||  ||  ||   ||  |  |  |   |||  ||  ||   ||  |||   |  |  |  |
Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 7
CS131783/c
LOCUS       CS131783               4377 bp    DNA     linear   PAT 02-AUG-2005
DEFINITION  Sequence 28 from Patent EP1555320.
ACCESSION   CS131783
VERSION     CS131783.1  GI:71794301
```

```
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Gerritse,G.
  TITLE     Expression system derived from the lipase regulation cascade of
            pseudomonas alcaligenes
  JOURNAL   Patent: EP 1555320-A 28 20-JUL-2005;
            Genencor International, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .4377
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             17.4%;  Score 174.2;  DB 6;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 0.0024;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||| || |||| |||| ||||     |  | || ||||     |    ||||  ||
Db        632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |   ||   |||| ||   ||| |||   |||  |    ||  |    |
Db        575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              | |  || ||| | |   | || ||                  | ||||||     |
Db        515 TGCTCCTCGTCCTGGGCGATGAAGTCC---------------TGCAGCAGCTCGCGCCAG 471

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||  |  ||| | |   ||  |||  ||||| |  |   || |   |  ||||  |||
Db        470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||  | |  |||||     |     |   |||   | |||  |  ||  || ||||
Db        410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || ||  | |   ||   |  |||  |  | || |||||    ||||||||    |||
Db        350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  ||| |||| || |     | | |||||  | |||  |   |||  || |||
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              |||     | |      |||| |       |  | |   | | || ||   |||   |||
Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              || | |  || | |||||         || || |||   |  |  || ||  ||||
Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | | ||  ||  ||| ||   |||||||   |  | ||  || |||  | ||||  |||
Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
              || ||  ||   || | | |   ||| ||| ||   || |||  | |  | |
Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5
```

```
RESULT 8
AB045630/c
LOCUS       AB045630               6979 bp    DNA     linear   BCT 19-JUN-2001
DEFINITION  Pseudomonas aeruginosa genes for glutamate-tRNA synthetase, CbrA,
            CbrB, complete cds.
ACCESSION   AB045630
VERSION     AB045630.1  GI:12082248
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (sites)
  AUTHORS   Nishijyo,T., Haas,D. and Itoh,Y.
  TITLE     The CbrA-CbrB two-component regulatory system controls the
            utilization of multiple carbon and nitrogen sources in Pseudomonas
            aeruginosa
  JOURNAL   Mol. Microbiol. 40 (4), 917-931 (2001)
   PUBMED   11401699
REFERENCE   2  (bases 1 to 6979)
  AUTHORS   Itoh,Y. and Nishijyo,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Yoshifumi Itoh, National Food Research
            Institute, Applied Microbiology; Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8642, Japan (E-mail:yosifumi@nfri.affrc.go.jp,
            Tel:81-298-38-8075, Fax:81-298-38-7996)
FEATURES             Location/Qualifiers
     source          1. .6979
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            284. .1165
                     /gene="gltX"
     CDS             284. .1165
                     /gene="gltX"
                     /note="putative"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="glutamate-tRNA synthetase"
                     /protein_id="BAB20865.1"
                     /db_xref="GI:12082249"
                     /translation="MTSSYVGRFAPTPSGYLHFGSLVAAVASYLDARAVGGRWLVRME
                     DLDPPREVPGAQQAILETLERYGFEWDGAVERQSERFPAYAAVVEQLLRSGLAYACTC
                     SRKQLEGFAGIYPGFCRDAGHAREDAAIRLRVPELEYRFVDRVQGEVRQHLGREVGDF
                     VIQRRDGLYAYQLAVVLDDAWQGITDIVRGADLLDSTPRQLYLQELLGLSQPRYLHVP
                     LIVQPDGHKLGKSYRSPPLPAEQAAAPLTRALRALGQRPPAELAQASASEALAWGVAH
                     WDATRIPRCATLPEERL"
     gene            1395. .4346
                     /gene="cbrA"
     CDS             1395. .4346
                     /gene="cbrA"
                     /function="signal sensor and protein histidine kinase"
                     /note="KinA-NtrB family
                     histidine at 766, phosphrylation site
                     histidine kinase domain,491 to 983 amino acids
                     sensor membrane domain, 1 to 490 amino acids"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CbrA"
                     /protein_id="BAB20866.1"
                     /db_xref="GI:12082250"
                     /translation="MLTSFSLTQLILISVTYLSTLFGIAWITERGYVPRRLVRHPLVY
                     TLSLGVYASAWAFYGTVGLAYQYGYGFLAIYLGISGAFLLAPVLLYPILRVTRAYQLS
                     SLADLFAFRFRSTWAGALTTLFMLIGVLPMLALQIQAVTDSISILTRDPAQERVAFVF
                     CTLITLFAILFGARHIATRERHEGLVFAIAFESLVKLVMLGSIGLYALYGVFGGPEGL
                     EVWLLQNQTALTTLHTPLAEGPWRTLLLVFFAAAIVMPHMFHMIFTENLNPRALISAS
                     WGLPLFLLLMSLAVPPILWAGLRLGASTTPEYYTIGLGLAVDSPALALAAFIGGISAA
                     SGLTIVMTLALSGMVLNHLVLPLYQPPAQGNIYRWLKWTRRLLIVAIIMASYAFYLLL
                     GAEQDLSNLGIVSFVATLQFLPGALSVLYWPTANRRGFISGLIAGMLVWAITMLLPLM
```

```
                     GNVQGLYLPLFDVIYVLDDSNWHLAALSSLAANVLVFTLVSLFTEASDEEKGAAEACA
                     VDNVRRPQRRELFAGSPQEFASQLAKPLGAKTAQKEVEQALRDLHLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDIVETFLPYKTSEESYVTEDIHFIESRLEDYHSRLTGLA
                     AELDTLRRYHRQTLQDLPMGVCSLAKDQEVLMWNRAIEELTGVGAQKVVGSRLSALPE
                     PWKGLLERFIDAPDEHLHKQRLVYDGHTRWLNLHKAAIEEPLAPGNSGLVLLVEDLTE
                     TQLLEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGELTEISEQ
                     ILDQTKRVSRIVQSLMSFAHSGSHLQALEPVCLSEVAQEAIGLLSLNRRSVEVEFFNL
                     CDPAHWVEGDSQRLAQVLINLLSNARDASPPGGAIRVRSEASEHTVDLVVEDEGSGIP
                     KAIMDQLFEPFFTTKDPGKGTGLGLALVYSIVEEHYGQITIESPTDHQREGGTRFRVT
                     LPRHPGPTAEL"
     gene            4373. .5809
                     /gene="cbrB"
     CDS             4373. .5809
                     /gene="cbrB"
                     /function="response regulator"
                     /note="Aspertate at 52 amino acids, phosphorylation site
                     NtrC family"
                     /codon_start=1
                     /transl_table=11
                     /product="CbrB"
                     /protein_id="BAB20867.1"
                     /db_xref="GI:12082251"
                     /translation="MAHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERYTIP
                     SFDMVVSDLRLPGAPGTELIKLAEGIPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILKDRQENRSAAPVSANGGKAGGERGASPAVADGEIGIIGSCAPMQELY
                     SKIRKVAPTDSTVLIQGESGTGKELVARALHNLSKRAKAPLISVNCAAIPETLIESEL
                     FGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRVGSV
                     QSQKVDVRLIAATHRDLKTLAKTGQFREDLYYRLHVISLKLPALRERGNDVMEIARAF
                     LARQCTRMGRAPLSFAHDAEQAIRHYPWPGNVRELENAIERAVILSESQEIHADLLGI
                     DIELDDLEDDFAADLGLGPAGAAASNHEPTEDLSLEDYFQHFVLEHQDHMTETELARK
                     LGISRKCLWERRQRLGIPRRKSGASADS"
ORIGIN

  Query Match           16.9%;  Score 168.8;  DB 1;  Length 6979;
  Best Local Similarity  51.4%;  Pred. No. 0.0043;
  Matches  485;  Conservative    0;  Mismatches  432;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              |||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db       3724 CAGGCGATGCCGGTGATCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCTAGGCGGCCG 3665

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||    |||||||| ||| |||| | ||||    | || | |    |    ||||||
Db       3664 ATGGAAGCCAGGCGCTCGGAGTGGACCAGCTTGTCTTCCAACAGTTGGGTTTCGGTGAGG 3605

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              || ||||| || || |||| |||||     | | || ||||    | |     |||| ||
Db       3604 TCTTCCACCAGTAGCACCAGGCCGC---TGTTGCCCGGCGCCAGGGGTTCCTCGATGGCC 3548

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |  |||   ||| |    | ||| |    || ||          || |    |
Db       3547 GCCTTGTGCAGGTTCAGCCAGCGGGTGTGGCCGTCGTAGACCAGGCGCTGCTTGTGCAGG 3488

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              | || |||   || |   | |                ||  | | ||| ||    |
Db       3487 TGCTCGTCGGGGGCGTCGATGAA--------------ACGTTCCAGCAGGCCCTTCCAG 3443

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||  | ||||        ||  |||  ||||||| |   || |  || | |||  |||
Db       3442 GGTTCGGGCAAGGCCGACAGGCGCGAACCGACGACCTTCTGCGCGCCGACCCCGGTGAGT 3383

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||||  |   |||||    |     |    |    |||   | ||| || |||||| ||
Db       3382 TCCTCGATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGGGAGCAGACGCCC 3323

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || |     ||    |   ||  |  | || |||| |   |||||  ||     |||
Db       3322 ATCGGCAGGTCTTGCAGGGTCTGCCGGTGGTAGCGGCGCAAGGTGTCGAGTTCGGCAGCC 3263

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
                |||    ||||| |  |     |  | | ||||| | ||| |   |  |||  ||||||
Db       3262 AGGCCAGTCAGCCGGGAGTGATAGTCCTCCAGGCGGCTCTCGATGAAGTGGATGTCCTCG 3203

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||    ||      ||| |      |    |  | || || |    ||
Db       3202 GTGACGTAGCTTTCTTCGCTGGTCTTGTAGGGAAGGAAGGTCTCCACGATGTC------- 3150

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              |   ||||     || || | ||||   || |||||    |  |  || ||  |||
Db       3149 --CTGGGCGACACTGGGTCCCATCAGCCCGGAAAGGTTCGCCTCGATGCGGTCGCGCAAG 3092

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
                | ||  |   ||| ||    ||||||   |  | || ||| |||  | ||||  ||
Db       3091 CGACGCAGGGCATAGGGGCGACGTTCGTCGAAGGGCAGGTGGAGATCGCGCAGGGCCTGC 3032

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||  ||   || | || |  ||| ||| ||   || | |  |     | | | |
Db       3031 TCGACCTCCTTCTGCGCGGTCTTGGCGCCCAGCGGCTTGGCCAACTGGCTGGCGAACTCC 2972

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              |          || |      ||||       ||  |||  |  | |     || |||
Db       2971 TGCGGCGAACCGGCGAACAGTTCGCGGCGTTGCGGACGGCGTACGTTGTCCACCGCGCAG 2912

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
                 ||| ||  |    |||||   |   | | |   | ||||||| || |    ||
Db       2911 GCCTCGGCGGCGCCTTTTTCCTCGTCGCTGGCCTCGGTGAACAGCGACACCAGGGTGAAT 2852

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
              |    ||||| | || |||||    | || ||  ||   |||
Db       2851 ACCAGCACGTTGGCCGCCAGCGAGGACAGCGCCGCCAGGTGCCA 2808


RESULT 9
AE004886/c
LOCUS       AE004886               10929 bp    DNA     linear   BCT 19-FEB-2003
DEFINITION  Pseudomonas aeruginosa PAO1, section 447 of 529 of the complete
            genome.
ACCESSION   AE004886 AE004091
VERSION     AE004886.1  GI:9950982
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa PAO1
  ORGANISM  Pseudomonas aeruginosa PAO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 10929)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  PUBMED    10984043
REFERENCE   2  (bases 1 to 10929)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 10929)
  AUTHORS   .
  CONSRTM   Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2003) Department of Molecular Biology and
            Biochemistry, Simon Fraser University, 8888 University Dr.,
            Burnaby, British Columbia V5A 1S6, Canada
COMMENT     ----------------------------------------------------------------
            This represents the February 3, 2003 version of the continually
```

```
        updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,
        from PseudoCAP (see http://www.pseudomonas.com for latest updates
        and links to alternate annotations). PseudoCAP is coordinated by
        Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
        E.W. Hancock (University of British Columbia, Canada). We welcome
        submission through www.pseudomonas.com of any proposed changes.

        'Protein name confidence' is used to rate our confidence of the
        accuracy of the protein name.
        Class 1: Function experimentally demonstrated in P. aeruginosa.
        Class 2: Function of highly similar gene experimentally
              demonstrated in another organism (and gene context consistent
              in terms of pathways its involved in, if known).
        Class 3: Function proposed based on presence of conserved amino
              acid motif, structural feature or limited sequence similarity
              to an experimentally studied gene.
        Class 4: Homologs of previously reported genes of unknown function,
              or no similarity to any previously reported sequences.
        ---------------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .10929
                     /organism="Pseudomonas aeruginosa PAO1"
                     /mol_type="genomic DNA"
                     /strain="PAO1"
                     /db_xref="taxon:208964"
     gene            127. .573
                     /gene="dksA"
                     /locus_tag="PA4723"
     CDS             127. .573
                     /gene="dksA"
                     /locus_tag="PA4723"
                     /note="Protein name confidence: Class 1 (function
                     experimentally demonstrated in P. aeruginosa)"
                     /codon_start=1
                     /transl_table=11
                     /product="suppressor protein DksA"
                     /protein_id="AAG08109.1"
                     /db_xref="GI:9950983"
                     /translation="MSTKAKQQSSQQMTRGFEPYQETKGEEYMSERMRAHFTAILNKW
                     KQELMEEVDRTVHHMQDEAANFPDPADRASQEEEFSLELRARDRERKLIKKIDETLQL
                     IEDEEYGWCDSCGVEIGIRRLEARPTATLCIDCKTLAEIREKQLGS"
     gene            641. .1522
                     /locus_tag="PA4724"
                     /note="synonym: yadB"
     CDS             641. .1522
                     /locus_tag="PA4724"
                     /note="Protein name confidence: Class 3 (function proposed
                     based on presence of conserved amino acid motif,
                     structural feature or limited sequence similarity to an
                     experimentally studied gene)"
                     /codon_start=1
                     /transl_table=11
                     /product="probable aminoacyl-transfer RNA synthetase
                     (class I)"
                     /protein_id="AAG08110.1"
                     /db_xref="GI:9950984"
                     /translation="MTSSYVGRFAPTPSGYLHFGSLVAAVASYLDARAVGGRWLVRME
                     DLDPPREVPGAQQAILETLERYGFEWDGAVERQSERFPAYAAVVEQLLRSGLAYACTC
                     SRKQLEGFAGIYPGPCRDAGHAREDAAIRLRVPELEYRFVDRVQGEVRQHLGREVGDF
                     VIQRRDGLYAYQLAVVLDDAWQGITDIVRGADLLDSTPRQLYLQELLGLSQPRYLHVP
                     LIVQPDGHKLGKSYRSPPLPAEQAAAPLTRALRALGQRPPAELAQASASEALAWGVAH
                     WDATRIPRCATLPEERL"
     gene            1752. .4703
                     /gene="cbrA"
                     /locus_tag="PA4725"
     CDS             1752. .4703
                     /gene="cbrA"
                     /locus_tag="PA4725"
                     /note="Protein name confidence: Class 1 (function
                     experimentally demonstrated in P. aeruginosa)"
                     /codon_start=1
                     /transl_table=11
                     /product="two-component sensor CbrA"
                     /protein_id="AAG08111.1"
                     /db_xref="GI:9950985"
                     /translation="MLTSFSLTQLILISVTYLSTLFGIAWITERGYVPRRLVRHPLVY
                     TLSLGVYASAWAFYGTVGLAYQYGYGFLAIYLGISGAFLLAPVLLYPILRVTRAYQLS
                     SLADLFAFRFRSTWAGALTTLFMLIGVLPMLALQIQAVTDSISILTRDPAQERVAFVF
                     CTLITLFAILFGARHIATRERHEGLVFAIAFESLVKLVMLGSIGLYALYGVFGGPEGL
                     EVWLLQNQTALTTLHTPLAEGPWRTLLLVFFAAAIVMPHMFHMIFTENLNPRALISAS
                     WGLPLFLLLMSLAVPPILWAGLRLGASTTPEYYTIGLGLAVDSPALALAAFIGGISAA
                     SGLTIVMTLALSGMVLNHLVLPLYQPPAQGNIYRWLKWTRRLLIVAIIMASYAFYLLL
                     GAEQDLSNLGIVSFVATLQFLPGALSVLYWPTANRRGFISGLIAGMLVWAITMLLPLM
                     GNVQGLYLPLFDVIYVLDDSNWHLAALSSLAANVLVFTLVSLFTEASDEEKGAAEACA
                     VDNVRRPQRRELFAGSPQEFASQLAKPLGAKTAQKEVEQALRDLHLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDIVETFLPYKTSEESYVTEDIHFIESRLEDYHSRLTGLA
                     AELDTLRRYHRQTLQDLPMGVCSLAKDQEVLMWNRAIEELTGVGAQKVVGSRLSALPE
                     PWKGLLERFIDAPDEHLHKQRLVYDGHTRWLNLHKAAIEEPLAPGNSGLVLLVEDLTE
                     TQLLEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGELTEISEQ
                     ILDQTKRVSRIVQSLMSFAHSGSHLQALEPVCLSEVAQEAIGLLSLNRRSVEVEFFNL
                     CDPAHWVEGDSQRLAQVLINLLSNARDASPPGGAIRVRSEASEHTVDLVVEDEGSGIP
                     KAIMDQLFEPFFTTKDPGKGTGLGLALVYSIVEEHYGQITIESPTDHQREGGTRFRVT
                     LPRHPGPTAEL"
     gene            4730. .6166
                     /gene="cbrB"
                     /locus_tag="PA4726"
     CDS             4730. .6166
                     /gene="cbrB"
                     /locus_tag="PA4726"
                     /note="Protein name confidence: Class 1 (function
                     experimentally demonstrated in P. aeruginosa)"
                     /codon_start=1
                     /transl_table=11
                     /product="two-component response regulator CbrB"
                     /protein_id="AAG08112.1"
                     /db_xref="GI:9950986"
                     /translation="MAHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERYTIP
                     SFDMVVSDLRLPGAPGTELIKLAEGIPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILKDRQENRSAAPVSANGGKAGGERGASPAVADGEIGIIGSCAPMQELY
                     SKIRKVAPTDSTVLIQGESGTGKELVARALHNLSKRAKAPLISVNCAAIPETLIESEL
                     FGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRVGSV
                     QSQKVDVRLIAATHRDLKTLAKTGQFREDLYYRLHVISLKLPALRERGNDVMEIARAF
                     LARQCTRMGRAPLSFAHDAEQAIRHYPWPGNVRELENAIERAVILSESQEIHADLLGI
                     DIELDDLEDDFAADLGLGPAGAAASNHEPTEDLSLEDYFQHFVLEHQDHMTETELARK
                     LGISRKCLWERRQRLGIPRRKSGASADS"
     gene            7067. .8470
                     /gene="pcnB"
                     /locus_tag="PA4727"
     CDS             7067. .8470
                     /gene="pcnB"
                     /locus_tag="PA4727"
                     /note="Protein name confidence: Class 2 (high similarity
                     to functionally studied protein); Alt protein name:
                     polynucleotide adenylyltransferase, plasmid copy number
                     protein"
                     /codon_start=1
                     /transl_table=11
                     /product="poly(A) polymerase"
                     /protein_id="AAG08113.1"
                     /db_xref="GI:9950987"
                     /translation="MLKKLIQTIRSPLRRPRAVRTTPEVIGNNQHSLRRDQFSRNAVK
                     VVEILQRAGYQAYVVGGCVRDQMLGIAPKDFDVATSATPEQVRAEFRNARIIGRRFKL
                     VHVHFGREIIEVATFRAHHPEGQDDGDSRASSNESGRILRDNVYGSLEDDAQRRDFTI
                     NALYFDVTSERLLDYANGVHDIRNRLIRLIGDPEQRYLEDPVRMLRAVRFAAKLDFEI
                     EKHSAAPIRRLAPLLREIPSARLFDEVLKLFLAGRAERTFELLVEYELFAPLFPASAK
                     ALQANPDYTGKLIRQALANTDARIRQGKPVTPAFLFAALLWPALPARVAQLQEKGMPA
                     IPAMQEAAHELISEQCQRIAIPKRFTLPIREIWDMQERLPRRQGKRADLLLENPRFRA
                     GYDFLLLRESAGEETEGLGQWWTDYQEVSDSERRNMIRDLVSQEDGSAPRKRRRGGNS
                     GRRRRGPRKEGSGGSGE"
     gene            8467. .8955
                     /gene="folK"
                     /locus_tag="PA4728"
     CDS             8467. .8955
                     /gene="folK"
                     /locus_tag="PA4728"
                     /note="Protein name confidence: Class 2 (high similarity
```

```
                to functionally studied protein); Alt protein name:
                6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase,
                7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase"
                /codon_start=1
                /transl_table=11
                /product="2-amino-4-hydroxy-6-
                hydroxymethyldihydropteridine pyrophosphokinase"
                /protein_id="AAG08114.1"
                /db_xref="GI:9950988"
                /translation="MIQRVYVALGSNLAEPREQIQAALDAFERLPETRLVAVSPLYIS
                DPLGPADQPRFVNGVAALDTNLAPLDLLDALQAIELEQGRVRDLRWGPRTLDLDILLF
                GEQLLDLPRLKVPHYHMQARAFVLYPLADLAPDLRLPDGRHLPELLAACPFEGIERLP

 Query Match             16.9%;  Score 168.8;  DB 1;  Length 10929;
 Best Local Similarity   51.4%;  Pred. No. 0.0034;
 Matches  485;  Conservative    0;  Mismatches  432;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||| ||| |||||| |||||||||||| || ||||| || || |||
Db       4081 CAGGCGATGCCGGTGATCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCTAGGCGGCCG 4022

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||    |||||||| |||  |||| | ||||   | || | |     |    |||||| |
Db       4021 ATGGAAGCCAGGCGCTCGGAGTGGACCAGCTTGTCTTCCAACAGTTGGGTTTCGGTGAGG 3962

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              || ||||| || || |||| |||||     | | || ||||   | |    ||||  ||
Db       3961 TCTTCCACCAGTAGCACCAGGCCGC---TGTTGCCCGGCGCCAGGGGTTCCTCGATGGCC 3905

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |  |||   ||| |   | ||| |   || ||       ||  |   |
Db       3904 GCCTTGTGCAGGTTCAGCCAGCGGGTGTGGCCGTCGTAGACCAGGCGCTGCTTGTGCAGG 3845

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              | || |||   || |   | |                ||  | | |||  ||    |
Db       3844 TGCTCGTCGGGGGCGTCGATGAA--------------ACGTTCCAGCAGGCCCTTCCAG 3800

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||  | ||||         ||  |||  |||||||  |   || |  || | |||  |||
Db       3799 GGTTCGGGCAAGGCCGACAGGCGCGAACCGACGACCTTCTGCGCGCCGACCCCGGTGAGT 3740

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||||  |   |||||    |     |   |   |||   | ||| || |||||| ||
Db       3739 TCCTCGATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGGGAGCAGACGCCC 3680

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || |    ||   |   ||  |  | || |||| |   |||||  ||    |||
Db       3679 ATCGGCAGGTCTTGCAGGGTCTGCCGGTGGTAGCGGCGCAAGGTGTCGAGTTCGGCAGCC 3620

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              |||    ||||| |  |     |  | | ||||| | ||| |   |  |||  ||||||
Db       3619 AGGCCAGTCAGCCGGGAGTGATAGTCCTCCAGGCGGCTCTCGATGAAGTGGATGTCCTCG 3560

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||    ||    ||| |      |   |  | || || |    ||
Db       3559 GTGACGTAGCTTTCTTCGCTGGTCTTGTAGGGAAGGAAGGTCTCCACGATGTC------- 3507

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              |   ||||     || || | ||||   || |||||     |  |  || ||  |||
Db       3506 --CTGGGCGACACTGGGTCCCATCAGCCCGGAAAGGTTCGCCTCGATGCGGTCGCGCAAG 3449

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | ||  |   ||| ||    ||||||    |  | ||| ||| |||  | ||||  ||
Db       3448 CGACGCAGGGCATAGGGGCGACGTTCGTCGAAGGGCAGGTGGAGATCGCGCAGGGCCTGC 3389

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||  ||   || | || |  ||| ||| ||   || | |  |    | | | |
Db       3388 TCGACCTCCTTCTGCGCGGTCTTGGCGCCCAGCGGCTTGGCCAACTGGCTGGCGAACTCC 3329

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              |         || |      ||||      ||  |||  |  | |     || |||
Db       3328 TGCGGCGAACCGGCGAACAGTTCGCGGCGTTGCGGACGGCGTACGTTGTCCACCGCGCAG 3269
```

```
Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              ||| ||  |    |||||   |   | |  |   | |||||| || |    ||
Db       3268 GCCTCGGCGGCGCCTTTTTCCTCGTCGCTGGCCTCGGTGAACAGCGACACCAGGGTGAAT 3209

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
                |    ||||| | || |||||     | || ||  ||   |||
Db       3208 ACCAGCACGTTGGCCGCCAGCGAGGACAGCGCCGCCAGGTGCCA 3165
```

```
RESULT 10
AE016791/c
LOCUS       AE016791              300933 bp    DNA     linear   BCT 12-DEC-2002
DEFINITION  Pseudomonas putida KT2440 section 18 of 21 of the complete genome.
ACCESSION   AE016791 AE015451
VERSION     AE016791.1  GI:26557036
KEYWORDS    .
SOURCE      Pseudomonas putida KT2440
  ORGANISM  Pseudomonas putida KT2440
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 300933)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Complete genome sequence and comparative analysis of the
            metabolically versatile Pseudomonas putida KT2440
  JOURNAL   Environ. Microbiol. 4 (12), 799-808 (2002)
REFERENCE   2  (bases 1 to 300933)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .300933
                     /organism="Pseudomonas putida KT2440"
                     /mol_type="genomic DNA"
                     /strain="KT2440"
                     /db_xref="taxon:160488"
     gene            153. .740
                     /locus_tag="PP4517"
     CDS             153. .740
                     /locus_tag="PP4517"
                     /note="identified by match to PFAM protein family HMM
                     PF03695"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAN70091.1"
                     /db_xref="GI:24986250"
                     /translation="MSFAEQLTRLQAFLDADELHEEALDYVAAHGYLTALSINAEDVP
                     EREWIDALFAEEPHYASEAQRTEIEATLVALKAHIARQLASDEEFDLPCDLDLTDEPD
                     DSDLRGWCIGFMEGVFLREEAWFENAEEEVSEMLLPIMVGSGLFDEQPEFADIASNAN
                     LQDDMIVQIPEALSALFLLLHAPDEKPALLKPRHH"
     gene            748. .1113
                     /locus_tag="PP4518"
     CDS             748. .1113
                     /locus_tag="PP4518"
                     /note="identified by match to PFAM protein family HMM
                     PF04304"
                     /codon_start=1
                     /transl_table=11
```

/product="conserved hypothetical protein"
/protein_id="AAN70092.1"
/db_xref="GI:24986251"
/translation="MRYLLLATGWLSVALGVLGIFLPVLPTTPFLLLAAACFARSSPR
FHHWLVHHPKLGPWIRDYLSGEGIPLKGKVYAIGLMWASIGLSCYLVPLFWARTFMLT
SAALVTLYILSLKTLHKPK"
gene 1425. .2783
/gene="tolC"
/locus_tag="PP4519"
CDS 1425. .2783
/gene="tolC"
/locus_tag="PP4519"
/note="similar to GB:X12387, SP:P05184, SP:P08684,
SP:P20815, PID:181354, PID:181372, PID:181374, PID:219963,
PID:35911, and PID:510086; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="agglutination protein"
/protein_id="AAN70093.1"
/db_xref="GI:24986253"
/translation="MRVLNPITSALLLALASANVQAMSITEAVQSAVDYHPQVSSNRN
SKLSADEDVKFARGGYYPSVDLVAGYGRQRSDNATTRAEGNHNKETLNYTQSELRLRQ
MIFDGFNTSNEVGRTEAVSTSRAYYTQAVAQDVALRAVEVYLEVLKRRELVTLAKNNL
QAHLRVNDQIGLRNERGVGSTADLDQSRARRALAENNLDTAEVDLADAEANFFSVIGR
APDELESPQSIKGEVPGTLEEARDGMRQNNPYIKSAQADVNAAEKQYEVGKSTFYPRF
DAILATGANNNTGGEKGHNNNDWQAGVEMNYNLFRGGSDKARLQSDAHKINQALDIRN
NALRELTENLSLAWNAMNNASKQLPTAREYAETTKRVRAAYQDQFGLGQRTLLDVLDS
ENELYNADRRYTEVRYTEEFSRYRVLATMGELLSKQHISLPPEALATTEVRTEARLPE
MR"
gene complement(2846. .3445)
/locus_tag="PP4520"
CDS complement(2846. .3445)
/locus_tag="PP4520"
/note="identified by match to PFAM protein family HMM
PF03942"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN70094.1"
/db_xref="GI:24986254"
/translation="MPRPRCERCQRPLDHCLCPLIPALDSRTRVILLQHPSETAHALN
TARLAALGLNNAELRIGEVFDDLHDLLATPGYHPVLLFPGDDAQVLTAYGEAGDKPLL
LIVPDGTWRKARKLLYLNPLLGGLPRVTLGQVAASRYRLRKAPEAGALSTIEAVVGAL
NTLEQPACFDALLAPFETLIEGQIRAMGVETFQRNHGRE"
gene complement(3510. .5075)
/locus_tag="PP4521"
CDS complement(3510. .5075)
/locus_tag="PP4521"
/note="similar to GP:9947522; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="aerotaxis receptor, putative"
/protein_id="AAN70095.1"
/db_xref="GI:24986255"
/translation="MRMNLPVTEHERTFPGDQRLISTTDLDSRITYCNDAFVAISGFT
YDELVGQPHNLVRHPDMPPAVFGHMWETIKQGKPWMGIVKNRAKNGDYYWVSAYVTAI
YEQGRISGYESVRSVPTREQIRRAEALYARLRDGRSPVPWAARLGHGLGHGWPLIGAG
LLSAAGYLWLPPYAALAVLMASLLLAWYWVEHRQNQAIRRTLAEHPKAFTSPLVALTY
SDNPGLKGQLDLAIISEEARLQTALTRLVDAGVGVKSRAAQSSDLSDAQAQMLDRQRS
ETDQSATAIAQMAATIQQVTHNVQSTAHAASDADQLAQQGSELALKSLKDMGSMSDAV
NDIGQAVNALAEQTQSIGSVVDVITSIAEQTNLLALNAAIEAARAGEQGRGFAVVADE
VRSLAQRTRASTEEIHHIIASLRAGAERAVSTASRGEQISRDSVHSVEAVQAALSGIA
VAVSRITGMSQQMATASEQQSHVAEDINQQIVRIAQLCDQSAGQARQGAEISQDLERM
AEYLHSLAERFNR"
gene complement(5214. .6107)
/locus_tag="PP4522"
CDS complement(5214. .6107)
/locus_tag="PP4522"
/note="similar to GP:9947370; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, LysR family"
/protein_id="AAN70096.1"
/db_xref="GI:24986256"
/translation="MASTLPDLKLLRIFISVVRHQGFANAQRELNLSTSAISTYMSQL
EGALGIVLCHRGRGGFSLTSKGELFHQETLRLLAELDGFEQYAAALKGELRGTLNLGV
IDSTVGDRALPLAEAIGAYSQEHPAVHLHLSVSSPYELQLGVQDNRLDLAIGAFSSRM
SGLLYQPLYREQHWLYCSSRHPLYAERRIPEQVVTQQRMVGRGYWSQAELARHGFKHS
AATVESMEAQLILILSGAYIGYLPEHYAQAWVDKGDLRVLSPSTFGYQAPFSLIIRRG
RSREPLIQTFRDLLKSQLNVG"
gene 6320. .7282
/locus_tag="PP4523"
CDS 6320. .7282
/locus_tag="PP4523"
/note="similar to SP:P37819; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="agmatinase, putative"
/protein_id="AAN70097.1"
/db_xref="GI:24986257"
/translation="MRPTVDKTLHQPLGGNEMPRFGGIATMLRLPHLQSAKGLDAAFI
GVPLDIGTSLRSGTRFGPRQIRAESVMIRPYNMATGAAPFDSLSVADIGDVAINTFNL
LDAVRIIEEAYDEIVEHNVIPMTLGGDHTITLPILRALHKKHGKIGLVHIDAHADVND
HMFGEKIAHGTTFRRAVEEGLLDCDRVVQIGLRAQGYTADDFNWSRRQGFRVVQAEEC
WHKSLEPLMAEVREKVGGGPVYLSFDIDGIDPAWAPGTGTPEIGGLTTIQAMEIIRGC
HGLDLIGCDLVEVSPPYDTTGNTSLLGANLLFEMLCVLPGVVRR"
gene 7415. .8794
/locus_tag="PP4524"
CDS 7415. .8794
/locus_tag="PP4524"
/note="identified by match to PFAM protein family HMM
PF04279"
/codon_start=1
/transl_table=11
/product="sodium-solute symporter, putative"
/protein_id="AAN70098.1"
/db_xref="GI:24986258"
/translation="MALDIIVVMIYTAGMLGLGWYGMRRAKTHEDYLVAGRNLGPVLY
MGTMATTVLGGASTVGTVRLGYVHGISGFWLCAALGLGIIAINLFLAKPLLRLRIFTV
TQVLEQRYNPTARQASAVIMLAYALMIGVTSTLAMATVLQVLLDLPFWASLMLGGGVV
VLYSTIGGMWSLTLTDIVQFVIKTVGLMFILLPVCLYKAGGWDTLVAKLPAASFQLTT
IGWDTIITYFLIYFFGILIGQDIWQRVFTARDEKVCQRAGTAAGVYCVLYGLACAAIG
MAAHVLMPDLANPNNAFAEMIKTTLPDGIRGLLMAAALAAMMSTASAGLLAASTTVTE
DLLPKLRGGKQSSLGTSRLFTLLTGLVVLGIALMVNDVINALTLAYNLLVGGMLIPLI
GAIFWKRATTAGAIASMSLGFATALLFMFLDGLEANTPIYYSLAVGLVSFVGVSLVSR
KSVAAVKLA"
gene complement(9174. .10067)
/locus_tag="PP4525"
CDS complement(9174. .10067)
/locus_tag="PP4525"
/note="similar to GP:9947361; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, LysR family"
/protein_id="AAN70099.1"
/db_xref="GI:24986259"
/translation="MEFSQLRIFQAVAEEGSVTRAAERLHRVPSNLSTRLRQLEEQLG
VELFLRERQRLQLSPAGKVLLDYANRMSALRDEALGAVRGGQPAGDFVLGTMYSTAAT
HLPALLARYHQSYPQVNLQVRAAPSGELLEGLLNHTLDAALVDGPPSLAGLDGVPLCD
EQLVLITSPEHPAVLTAKDVAGKAVFTFRQGCSYRMRLEAWYAHAHTPMGRVMEIESY
QSMLACVIAGAGVAMMAQSMLDSLPGRDRVRVHSLQAPFDQAVTWLMWRQGMRGANLQ
AWIDLQQSETINQSSECAATA"
gene 10172. .11344

Query Match 16.7%; Score 166.8; DB 1; Length 300933;
Best Local Similarity 52.2%; Pred. No. 0.00097;
Matches 484; Conservative 0; Mismatches 417; Indels 27; Gaps 4;

Qy 1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
||||||||||||||| |||||||||| |||||||||||||| ||||| || || |||
Db 204406 CAGGCGATGCCGGTGATCGGGTTGCCGATCTCGTGGGCCACGCCGGCGGCCAGGCGGCCG 204347

```
Qy          61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
               ||   |||||||||| |||  |||  | ||||    | || || ||  |    |||||  |
Db      204346 ATGCTGGCCAGGCGCTCGGAGTGCACCAGCTTGTCTTCCAGTGCCTGGGTTTCGGTAAGG 204287

Qy         121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
               || ||||| ||||| |||| ||||       |||   | ||      ||   |||| ||
Db      204286 TCTTCCACCAGCAGTACCAGCCCGCTGTTACCGG---GGGCCAATGGCTCGTCGATGGCT 204230

Qy         181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
               |||  ||||| |  |||   ||||| | | ||| |||     ||        ||  |   |
Db      204229 GCCTTGTGCAGGTTCAGCCAGCGCGGCTGGCCGTCCAGCGCCAGGCGCTGCTTGTGCAGG 204170

Qy         241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
               | || ||| | ||      | |                ||     | |||  |     |
Db      204169 TGCTCGTCGGGCACGTTGATGAA--------------GCCTTGCAGCAGGCCGCGCCAA 204125

Qy         301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
               ||  |     |    | | ||  |||  ||||| |  |       |  ||  ||||  ||||
Db      204124 GGCTCGTCGATGGTGACCAGGCGCGAGCCGACCACATGCTTGGCGGCGATGCCGGTGAGC 204065

Qy         361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
               | || | |  |  |||    |   ||| |   |||   | |||  | ||||||  ||
Db      204064 TCTTCCATGGCCTTGTTCCACATCAGGATTTCCTGGTCCTTGGCCAGCGAGCAGACGCCC 204005

Qy         421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
               |  || | || |   ||   |  ||| |  | || |||| |  |||||||||    |||
Db      204004 ATGGGCAACTCCTGCAGGGTCTGGCGATGGTAGCGGCGCAAGGCATCGAGCTCGGCGGCC 203945

Qy         481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
               |||   || |||| | |  |    | ||||| |  || |   |  |||  || |||
Db      203944 AGCCCGGTCAGGCGCGAGTGGTAGTCTTCCAGGCGGCTTTCGATGAAGTGGATGTCTTCG 203885

Qy         541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
               ||   | ||  |   ||  |||||||  |  |||| |   | |||  ||    |
Db      203884 GT---CACGTAGTTTTCGTTGCCGGACTTGTACGGCAGGAAAGTCTCGACCATGTC---- 203832

Qy         601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
               |   || |   ||| || | |||    |||| ||||   |  |  || ||  ||||
Db      203831 --CTGGGCCACGCTCGGCCCCATCAGGCCGGACAAGTTGGCTTCGATGCGGTCGCGCAGG 203774

Qy         661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
               | | ||  |   ||| ||    ||||||  ||  | || ||| ||| |  | ||  ||
Db      203773 CGGCGCAGGGCATAGGGGCGGCGTTCGTCGAACGGCAGGTAGAGGTCGCGCAGTGCTTGC 203714

Qy         721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
               ||| ||  |   || |||| |  ||| ||| ||   || |||  |   | |   |
Db      203713 TCGACTTCCTTTTGCGCGGCCTTGGCGCCCAGCGGCTTGGCCAGCTGGGTGGCGAATTCC 203654

Qy         781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
               |      | ||      | ||||    |   | | |||  |  |         |  ||
Db      203653 TGCGGCGAGGCGGCGTGCAGCTCGCGGCGCTGTGGGCGGCGCACGTTATCCACGGCACAG 203594

Qy         841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
               ||||||   |  ||| ||  ||   |    |   | ||| |||| |    || |
Db      203593 GCTTCGGCCGCACTGACCTCTTCGGTGCTGGCGTTGGTGAACAGTGAGATCAGGGTGAAC 203534

Qy         901 GTCGCGCCGTTGACGGCGAGCGACACGA 928
               |   ||||| | || | |||  |||
Db      203533 AGCAGCACGTTGGCCGCCAACGAGGCGA 203506


RESULT 11
CP000076_60/c
WPCOMMENT
  Sequence split into 71 fragments  LOCUS CP000076 Accession CP000076
     Fragment Name      Begin     End
     CP000076_00            1    110000
     CP000076_01       100001    210000
     CP000076_02       200001    310000
     CP000076_03       300001    410000
     CP000076_04       400001    510000
     CP000076_05       500001    610000
     CP000076_06       600001    710000
     CP000076_07       700001    810000
     CP000076_08       800001    910000
     CP000076_09       900001   1010000
     CP000076_10      1000001   1110000
     CP000076_11      1100001   1210000
     CP000076_12      1200001   1310000
     CP000076_13      1300001   1410000
     CP000076_14      1400001   1510000
     CP000076_15      1500001   1610000
     CP000076_16      1600001   1710000
     CP000076_17      1700001   1810000
     CP000076_18      1800001   1910000
     CP000076_19      1900001   2010000
     CP000076_20      2000001   2110000
     CP000076_21      2100001   2210000
     CP000076_22      2200001   2310000
     CP000076_23      2300001   2410000
     CP000076_24      2400001   2510000
     CP000076_25      2500001   2610000
     CP000076_26      2600001   2710000
     CP000076_27      2700001   2810000
     CP000076_28      2800001   2910000
     CP000076_29      2900001   3010000
     CP000076_30      3000001   3110000
     CP000076_31      3100001   3210000
     CP000076_32      3200001   3310000
     CP000076_33      3300001   3410000
     CP000076_34      3400001   3510000
     CP000076_35      3500001   3610000
     CP000076_36      3600001   3710000
     CP000076_37      3700001   3810000
     CP000076_38      3800001   3910000
     CP000076_39      3900001   4010000
     CP000076_40      4000001   4110000
     CP000076_41      4100001   4210000
     CP000076_42      4200001   4310000
     CP000076_43      4300001   4410000
     CP000076_44      4400001   4510000
     CP000076_45      4500001   4610000
     CP000076_46      4600001   4710000
     CP000076_47      4700001   4810000
     CP000076_48      4800001   4910000
     CP000076_49      4900001   5010000
     CP000076_50      5000001   5110000
     CP000076_51      5100001   5210000
     CP000076_52      5200001   5310000
     CP000076_53      5300001   5410000
     CP000076_54      5400001   5510000
     CP000076_55      5500001   5610000
     CP000076_56      5600001   5710000
     CP000076_57      5700001   5810000
     CP000076_58      5800001   5910000
     CP000076_59      5900001   6010000
     CP000076_60      6000001   6110000
     CP000076_61      6100001   6210000
     CP000076_62      6200001   6310000
     CP000076_63      6300001   6410000
     CP000076_64      6400001   6510000
     CP000076_65      6500001   6610000
     CP000076_66      6600001   6710000
     CP000076_67      6700001   6810000
     CP000076_68      6800001   6910000
     CP000076_69      6900001   7010000
     CP000076_70      7000001   7074893
Continuation (61 of 71) of CP000076 from base 6000001 (CP000076 Pseudomonas fluorescens I

  Query Match           16.5%;  Score 165.6;  DB 1;  Length 110000;
  Best Local Similarity  51.2%;  Pred. No. 0.0019;
  Matches  483;  Conservative    0;  Mismatches  434;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||| | |||||||||||| ||| |||||||| || || || || ||||||
```

```
Db      38687 CAGGCGATGCCGGTAATCGGGTTGCCGATCTCATGGGCCACCCCGGCCGCCAGGCGCCCG 38628

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   ||||||||| |||  |||  | ||||   | || || |   |    |||||  |
Db      38627 ATGCTGGCCAGGCGTTCGGAGTGCACCAGCTTGTCTTCCAGGGTCTGGGTTTCGGTCAGG 38568

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              || ||||| ||||| |||| |||||     |  | || | ||      ||  |||||  |||
Db      38567 TCTTCCACCAGCAGTACCAGGCCGC---TGTTGCCCGGGGCCAGAGGCTCATCGATGGCG 38511

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |  |||   ||| | | ||  | | |  ||  |     |  |     |
Db      38510 GCCTTGTGCAGGTTCAGCCAGCGGGTCTGGCCATCGAGGGCCAGGTGTTGTTTGTGCAAG 38451

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              || | | |  |      || ||                  | ||| |||    ||
Db      38450 TGTTCGTCCGGCAGGTTGATAAAGCCC---------------TGCAGCAGGTCCTTCCAC 38406

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||   |  |    |  ||  | |  || || | ||   |  |       |||  ||||
Db      38405 GGGTTGGTAATGGTGCTCAGGCGTGAGCCCACCACGTGCTGGGCTGGTATGCCTGTGAGC 38346

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||  |  |  |  |||    | |  |||| |   |||   | ||| || |||||| ||
Db      38345 TCCTCCATGGCCTTGTTCCACATGAGGATCTCCTGGTCCTTGGCCAGGGAGCAGACTCCC 38286

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || ||||     ||        ||  |  | || |||||   |||||||||    |||
Db      38285 ATCGGCAGCTCTTGCAGGGTTTGCCGGTGATAGCGGCGCAGTGCATCGAGCTCGGCGGCC 38226

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  ||| || |  | | |  | | || || |  || |   |  |||  ||||||
Db      38225 AGGCCGGTGAGGCGTGAGTGGTAGTCCTCCAGCCGGCTTTCGATGAAGTGGATGTCCTCG 38166

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||    |     | ||| |     |  | |    | |  || ||   ||
Db      38165 GTGACGTAGTTTTCTCCACCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTC------- 38113

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              |   ||||     ||||| | || |   |||||||||   |  |  || ||  ||||
Db      38112 --CTGGGCGACGCTGGGGCCCATCAACCCGGACAGGTTGGCTTCGATGCGGTCGCGCAGG 38055

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | ||  ||   ||  ||     |||||     |  |  |  || ||   | ||||  ||
Db      38054 CGCCGCAGGGCGTAAGGTCGGCGTTCGTCAAAGGGCAGATACAGATCACGCAGGGCCTGC 37995

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              || ||  ||  ||| |||| |   || ||| ||   || |||  |    | | | |
Db      37994 TCCACTTCCTTCTGGGCTGCCTTGGCCCCCAGCGGCTTGGCCAGCTGGGTGGCGAACTCC 37935

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
                |  ||         | ||    |  ||   |||  |  | |       |  ||
Db      37934 TGGGGCGAGGCGGCGTGCAGTTCCCGGCGCTGCGGACGGCGCACGTTGTCCACGGCACAG 37875

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              ||| ||   |  ||||||  ||    |    |   | ||||||||  |    || |
Db      37874 GCTTCGGCGGCGCTGGCCTCCTCGGTGCTGGCATTGGTGAACAGCGAGATCAGGGTGAAC 37815

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
              ||  | ||||| | || |||||  |||  | ||  |   |||
Db      37814 ATCAGGACGTTGGCCGCCAGCGAGGCGATGGCCGCCATATGCCA 37771


RESULT 12
CP000075_09
WPCOMMENT
  Sequence split into 61 fragments  LOCUS CP000075 Accession CP000075
     Fragment Name      Begin     End
     CP000075_00            1    110000
     CP000075_01       100001    210000
     CP000075_02       200001    310000
     CP000075_03       300001    410000
     CP000075_04       400001    510000
     CP000075_05       500001    610000
     CP000075_06       600001    710000
     CP000075_07       700001    810000
     CP000075_08       800001    910000
     CP000075_09       900001   1010000
     CP000075_10      1000001   1110000
     CP000075_11      1100001   1210000
     CP000075_12      1200001   1310000
     CP000075_13      1300001   1410000
     CP000075_14      1400001   1510000
     CP000075_15      1500001   1610000
     CP000075_16      1600001   1710000
     CP000075_17      1700001   1810000
     CP000075_18      1800001   1910000
     CP000075_19      1900001   2010000
     CP000075_20      2000001   2110000
     CP000075_21      2100001   2210000
     CP000075_22      2200001   2310000
     CP000075_23      2300001   2410000
     CP000075_24      2400001   2510000
     CP000075_25      2500001   2610000
     CP000075_26      2600001   2710000
     CP000075_27      2700001   2810000
     CP000075_28      2800001   2910000
     CP000075_29      2900001   3010000
     CP000075_30      3000001   3110000
     CP000075_31      3100001   3210000
     CP000075_32      3200001   3310000
     CP000075_33      3300001   3410000
     CP000075_34      3400001   3510000
     CP000075_35      3500001   3610000
     CP000075_36      3600001   3710000
     CP000075_37      3700001   3810000
     CP000075_38      3800001   3910000
     CP000075_39      3900001   4010000
     CP000075_40      4000001   4110000
     CP000075_41      4100001   4210000
     CP000075_42      4200001   4310000
     CP000075_43      4300001   4410000
     CP000075_44      4400001   4510000
     CP000075_45      4500001   4610000
     CP000075_46      4600001   4710000
     CP000075_47      4700001   4810000
     CP000075_48      4800001   4910000
     CP000075_49      4900001   5010000
     CP000075_50      5000001   5110000
     CP000075_51      5100001   5210000
     CP000075_52      5200001   5310000
     CP000075_53      5300001   5410000
     CP000075_54      5400001   5510000
     CP000075_55      5500001   5610000
     CP000075_56      5600001   5710000
     CP000075_57      5700001   5810000
     CP000075_58      5800001   5910000
     CP000075_59      5900001   6010000
     CP000075_60      6000001   6093698
Continuation (10 of 61) of CP000075 from base 900001 (CP000075 Pseudomonas syringae pv.

  Query Match            16.0%;  Score 160.4;  DB 1;  Length 110000;
  Best Local Similarity  51.7%;  Pred. No. 0.004;
  Matches  480;  Conservative    0;  Mismatches  421;  Indels   27;  Gaps    4;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||| |||||||||| ||| || ||||  ||||| ||||| || ||||| || || |||
Db      43401 CAGGCAATGCCGGTGATCGGATTACCGATTTCGTGCGCCACACCGGCGGCCAGGCGGCCG 43460

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||     |||||||| ||   |||  | ||||   | ||||||     ||   |||||  |
Db      43461 ATGCTTGCCAGGCGTTCCGAGTGCACCAGCTTGTCTTCGAGCATCTGCGTGTCGGTCAGG 43520

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||| || ||||  ||||  |||||     |  | || ||||     ||   |||||  |
Db      43521 TCCTCGACCAGCACCACCAGGCCGC---TGTTGCCCGGCGCCAGAGGCTCGTTGATGGCT 43577
```

```
Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  |||||     ||   ||| | | | ||| |||      | ||    ||
Db      43578 GCCTTGTGCAGATTGAGCCAGCGTGTCTGGCCGTCCAGACCGAGCTTCTGCT-------- 43629

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                |  |     | |   || |  || ||      | ||  | |||   |    ||
Db      43630 -------TGTGCAAATGCTCATCGGGCAGGCTGGTAAAACCGGTCAGCAGGCTCTTCCAC 43682

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||  | |  |     || ||  |||  ||||| |      || |  ||  ||||   |
Db      43683 GGCTCGGCAATGGTTTCCAGTCGCGAGCCGACCACACGCTGCGCGACGATGCCGGTCAAT 43742

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              | || | |  || |||     |    |||| |    ||   | |||  | |||||| ||
Db      43743 TCTTCCATGGCCCTGTTCCACATCAGGATCTCCTGATCTTCGGCCAGCGAGCAGACGCCC 43802

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || |||| |   ||   |  |||  |  | || ||||||  |||||| ||    ||
Db      43803 ATCGGCAGCTCCTGTAGGGTCTGGCGGTGGTAGCGACGCAGGGCATCGAGTTCGGCGGCA 43862

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  ||| ||||  | |  |  | | || || |  || |  |  | |  || |||
Db      43863 AGGCCGGTGAGGCGCGAGTGGTAGTCCTCAAGCCGGCTTTCAATGAAATGTATGTCTTCG 43922

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||   ||        ||| ||| ||    |  |||        |  ||| ||    |
Db      43923 GTGA---CGTATTTTTCGCTGCCTGATTTGTACGGCAGAAAGGTTTCCACCATGTC---- 43975

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
                |   || |    ||| || | ||||   |||||||||    |  |  || ||  ||||
Db      43976 --CTGGGCAACGCTCGGCCCCATCAGCCCGGACAGGTTGGCCTCGATGCGGTCGCGCAGG 44033

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | | ||  |    || ||   |||||||   |  | || ||| |||  | | ||  ||
Db      44034 CGGCGCAGGGCATAAGGGCGGCGCTCGTCGAAGGGCAGGTAGAGATCGCGCAGCGCTTGC 44093

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||  ||   || || |    ||| ||| ||   || ||    |    | | | |
Db      44094 TCGACCTCCTTCTGCGCGGCTTTCGCGCCCAGCGGCTTGGCCAGTTGGGTGGCGAACTCC 44153

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              |      |  ||        | ||   |   | | | |  |  | |  |   | |||
Db      44154 TGCGGCGACACGGCGTGCAGTTCACGGCGCTGTGGGCGACGCACGTTGTCGACTGCGCAG 44213

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              |||  ||| ||   || ||   |   |   |  | ||||||||  |    || |
Db      44214 GCCTCGGCGGCCCCGGCTTCTTCAGGGCTGGCATTGGTGAACAGCGAGATCAGGGTGAAC 44273

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGA 928
              |  | |||||  ||||||||||  |||
Db      44274 AGCAAGACGTTGGCGGCGAGCGAGGCGA 44301
```

RESULT 13
AE016853_10
WPCOMMENT
Sequence split into 64 fragments LOCUS AE016853 Accession AE016853

Fragment Name	Begin	End
AE016853_00	1	110000
AE016853_01	100001	210000
AE016853_02	200001	310000
AE016853_03	300001	410000
AE016853_04	400001	510000
AE016853_05	500001	610000
AE016853_06	600001	710000
AE016853_07	700001	810000
AE016853_08	800001	910000
AE016853_09	900001	1010000
AE016853_10	1000001	1110000
AE016853_11	1100001	1210000
AE016853_12	1200001	1310000
AE016853_13	1300001	1410000
AE016853_14	1400001	1510000
AE016853_15	1500001	1610000
AE016853_16	1600001	1710000
AE016853_17	1700001	1810000
AE016853_18	1800001	1910000
AE016853_19	1900001	2010000
AE016853_20	2000001	2110000
AE016853_21	2100001	2210000
AE016853_22	2200001	2310000
AE016853_23	2300001	2410000
AE016853_24	2400001	2510000
AE016853_25	2500001	2610000
AE016853_26	2600001	2710000
AE016853_27	2700001	2810000
AE016853_28	2800001	2910000
AE016853_29	2900001	3010000
AE016853_30	3000001	3110000
AE016853_31	3100001	3210000
AE016853_32	3200001	3310000
AE016853_33	3300001	3410000
AE016853_34	3400001	3510000
AE016853_35	3500001	3610000
AE016853_36	3600001	3710000
AE016853_37	3700001	3810000
AE016853_38	3800001	3910000
AE016853_39	3900001	4010000
AE016853_40	4000001	4110000
AE016853_41	4100001	4210000
AE016853_42	4200001	4310000
AE016853_43	4300001	4410000
AE016853_44	4400001	4510000
AE016853_45	4500001	4610000
AE016853_46	4600001	4710000
AE016853_47	4700001	4810000
AE016853_48	4800001	4910000
AE016853_49	4900001	5010000
AE016853_50	5000001	5110000
AE016853_51	5100001	5210000
AE016853_52	5200001	5310000
AE016853_53	5300001	5410000
AE016853_54	5400001	5510000
AE016853_55	5500001	5610000
AE016853_56	5600001	5710000
AE016853_57	5700001	5810000
AE016853_58	5800001	5910000
AE016853_59	5900001	6010000
AE016853_60	6000001	6110000
AE016853_61	6100001	6210000
AE016853_62	6200001	6310000
AE016853_63	6300001	6397126

Continuation (11 of 64) of AE016853 from base 1000001 (AE016853 Pseudomonas syringae pv.

Query Match 15.5%; Score 155.6; DB 1; Length 110000;
Best Local Similarity 51.4%; Pred. No. 0.008;
Matches 477; Conservative 0; Mismatches 424; Indels 27; Gaps 4;

```
Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||| |||||||||| |||||||||| | |||||| |||||||| || || || || |||
Db      49234 CAGGCAATGCCGGTGATCGGGTTGCCAATCTCGTGCGCCACGCCTGCCGCCAGGCGGCCG 49293

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   ||||||||| ||    ||  | || |   | ||||||    |    |||||| |
Db      49294 ATGCTGGCCAGGCGTTCCGAATGCACCAGTTTGTCTTCGAGCATCTGAGTGTCGGTGAGG 49353

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              || || || |||| ||||  ||||     |  | |  |||||     ||   | ||  ||
Db      49354 TCTTCGACCAGCACGACCAGCCCGC---TGTTGCCTGGCGCGAGAGGCTCGTTGATGGCC 49410

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              ||   ||||| |  |||   ||| | | ||||| |||      | ||    ||  |
Db      49411 GCTTTGTGCAGGTTCAGCCAGCGGGTCTGCCCGTCCAGCCCGAGCTTCTGCTTGTGCAAA 49470

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
```

```
Db      49471 TGCTCATCGGGAAGGCTGGTGAA---------------ACCGGTCAGCAGGCCTTTCCAG 49515

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||  | |  |    ||| ||  |||  ||||| |      |  |   ||    |||   ||
Db      49516 GGTTCGGCAATGGTGTCCAGGCGCGAGCCGACCACCCGTTGGGCGACGATTCCGGTCAGT 49575

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              | || | |  ||||||     |     |||| |     ||    | |||   | ||||||  ||
Db      49576 TCTTCCATGGCCCGGTTCCACATCAGGATCTCCTGATCTTCGGCCAGCGAGCAGACGCCC 49635

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || ||||      |||  |    ||   |  | || |||| |    |||||  ||     |||
Db      49636 ATCGGTAGCTCTTGCAGCGTCTGACGGTGGTAGCGGCGCAAGGCGTCGAGTTCGGCGGCC 49695

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||   || ||||  | |  |  | | || || |  || |    |  |||  || |||
Db      49696 AGGCCGGTCAGGCGCGAGTGGTAGTCCTCAAGCCGGCTTTCGATGAAATGGATGTCTTCG 49755

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||   ||       | | ||| |||  |  ||||  |   |  ||  ||    |
Db      49756 GTGA---CGTATTTTTCACTGCCTGACTTGTACGGCAGGAAGGTTTCGACCATGTC---- 49808

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              |   ||||     || || | ||||   |||||||||   |  | ||| ||  ||||
Db      49809 --CTGGGCGACGCTTGGCCCCATCAGCCCGGACAGGTTGGCTTCGATCCGGTCACGCAGG 49866

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | | | | |    || ||   |||||||  ||  | || ||| |||  | | ||  |
Db      49867 CGGCGCAACGCATAAGGGCGGCGCTCGTCGAACGGCAGGTAGAGGTCGCGCAGCGCCTGT 49926

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||  |    || || |    ||| | | ||   || ||   |    | | | |
Db      49927 TCGACCTCTTTTTGTGCGGCTTTCGCGCCCAACGGCTTGGCCAGTTGGGTGGCGAACTCC 49986

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
                  ||      | ||||   |  || | | |  || | |   |  | |||
Db      49987 TGAGGCGAAACGGCATGCAGCTCGCGGCGCTGCGGGCGACGCATGTTGTCGACGGCGCAG 50046

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              |||  ||   |  || ||    |    |    | |||||||  |     | ||
Db      50047 GCCTCGGCGGCGCTGACTTCTTCGGGGCTGGCGTTGGTGAACAGCGAGATCAGGGTAAAC 50106

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGA 928
              |    ||||| | |||||||| |||
Db      50107 AGCAGCACGTTGGCTGCGAGCGACGCGA 50134


RESULT 14
CP000058_10
WPCOMMENT
  Sequence split into 60 fragments  LOCUS CP000058 Accession CP000058
     Fragment Name      Begin     End
     CP000058_00            1   110000
     CP000058_01       100001   210000
     CP000058_02       200001   310000
     CP000058_03       300001   410000
     CP000058_04       400001   510000
     CP000058_05       500001   610000
     CP000058_06       600001   710000
     CP000058_07       700001   810000
     CP000058_08       800001   910000
     CP000058_09       900001  1010000
     CP000058_10      1000001  1110000
     CP000058_11      1100001  1210000
     CP000058_12      1200001  1310000
     CP000058_13      1300001  1410000
     CP000058_14      1400001  1510000
     CP000058_15      1500001  1610000
     CP000058_16      1600001  1710000
     CP000058_17      1700001  1810000
     CP000058_18      1800001  1910000
     CP000058_19      1900001  2010000
     CP000058_20      2000001  2110000
     CP000058_21      2100001  2210000
     CP000058_22      2200001  2310000
     CP000058_23      2300001  2410000
     CP000058_24      2400001  2510000
     CP000058_25      2500001  2610000
     CP000058_26      2600001  2710000
     CP000058_27      2700001  2810000
     CP000058_28      2800001  2910000
     CP000058_29      2900001  3010000
     CP000058_30      3000001  3110000
     CP000058_31      3100001  3210000
     CP000058_32      3200001  3310000
     CP000058_33      3300001  3410000
     CP000058_34      3400001  3510000
     CP000058_35      3500001  3610000
     CP000058_36      3600001  3710000
     CP000058_37      3700001  3810000
     CP000058_38      3800001  3910000
     CP000058_39      3900001  4010000
     CP000058_40      4000001  4110000
     CP000058_41      4100001  4210000
     CP000058_42      4200001  4310000
     CP000058_43      4300001  4410000
     CP000058_44      4400001  4510000
     CP000058_45      4500001  4610000
     CP000058_46      4600001  4710000
     CP000058_47      4700001  4810000
     CP000058_48      4800001  4910000
     CP000058_49      4900001  5010000
     CP000058_50      5000001  5110000
     CP000058_51      5100001  5210000
     CP000058_52      5200001  5310000
     CP000058_53      5300001  5410000
     CP000058_54      5400001  5510000
     CP000058_55      5500001  5610000
     CP000058_56      5600001  5710000
     CP000058_57      5700001  5810000
     CP000058_58      5800001  5910000
     CP000058_59      5900001  5928787
Continuation (11 of 60) of CP000058 from base 1000001 (CP000058 Pseudomonas syringae pv.

  Query Match           13.8%;  Score 138;  DB 1;  Length 110000;
  Best Local Similarity  50.2%;  Pred. No. 0.1;
  Matches  466;  Conservative    0;  Mismatches  435;  Indels   27;  Gaps    4;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||| |||||||||| ||||||||||  ||||| |||||||| || || || || ||
Db      27259 CAGGCAATGCCGGTGATCGGGTTGCCGATTTCGTGCGCCACGCCTGCTGCCAGTCGACCA 27318

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || ||   |||  | ||||   | || |||    ||   |||||  |
Db      27319 ATGCTGGCCAGCCGCTCTGAGTGCACCAGCTTGTCTTCAAGCATCTGCGTGTCGGTCAGG 27378

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||| || ||||  |||| ||| |     |  | |  | |||     ||   | ||  ||
Db      27379 TCCTCGACCAGCACCACCAGGCCAC---TGTTGCCAGGGGCGAGAGGCTCGTTGATGGCC 27435

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |   ||    ||| | | | ||| |||      |  |    ||
Db      27436 GCCTTGTGCAGGTTTAGCCAGCGGGTCTGGCCGTCCAGACCGAGTTTCTGCT-------- 27487

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                    |  |  | | |   || |  || |||   ||  ||  | |||         |
Db      27488 -------TGTGCAGGTGCTCATCGGGCAGGCTGGCAAAGCCGGTCAGCAGGCTTTTCCAG 27540

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||  | | |      || ||  |||  ||||| | |     |  |  ||  ||||   |||
Db      27541 GGTTCGGCAATGGTTTCCAGTCGCGAACCGACCACGCGCTGGGCGACGATGCCGGTCAGC 27600

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              | || |    || |||     |    |||| |    ||    | |||  | ||| || ||
Db      27601 TCTTCCATCGCCCTGTTCCACATCAGGATCTCCTGATCTTCGGCCAGCGAGCACACACCC 27660
```

```
Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
Db      27661 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGTAGCGGCGCAGGGCGTCGAGTTCGGCAGCA 27720

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
Db      27721 AGGCCGGTCAGGCGCGAGTGGTAGTCCTCAAGACGGCTTTCAATAAAATGTATGTCTTCG 27780

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
Db      27781 GTGA---CATATTTTTCGCTGCCCGATTTGTACGGCAGGAAGGTTTCCACCATGTC---- 27833

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
Db      27834 --CTGGGCGACACTTGGCCCCATCAACCCGGACAGGTTGGCCTCGATCCTGTCGCGCAGG 27891

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
Db      27892 CGGCGCAGGGCATAAGGACGGCGCTCGTCGAATGGCAGGTACAGATCGCGCAGCGCCTGT 27951

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
Db      27952 TCGACTTCCTTCTGTGCGGCTTTCGCGCCCAGCGGCTTGGCCAGTTGGGTGGCGAATTCC 28011

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
Db      28012 TGGGGCGACACGGCATGCAGTTCACGGCGTTGCGGGCGGCGCACGTTATCGACCGCGCAG 28071

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
Db      28072 GCCTCGGCTGCACCGACTTCTTCAGGGCTGGCATTGGTGAACAGGGATATCAGGGTGAAC 28131

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGA 928
Db      28132 AACAACACGTTGGCCGCGAGGGACGCGA 28159
```

RESULT 15
CP000011_08/c
WPCOMMENT
Sequence split into 24 fragments LOCUS CP000011 Accession CP000011

Fragment Name	Begin	End
CP000011_00	1	110000
CP000011_01	100001	210000
CP000011_02	200001	310000
CP000011_03	300001	410000
CP000011_04	400001	510000
CP000011_05	500001	610000
CP000011_06	600001	710000
CP000011_07	700001	810000
CP000011_08	800001	910000
CP000011_09	900001	1010000
CP000011_10	1000001	1110000
CP000011_11	1100001	1210000
CP000011_12	1200001	1310000
CP000011_13	1300001	1410000
CP000011_14	1400001	1510000
CP000011_15	1500001	1610000
CP000011_16	1600001	1710000
CP000011_17	1700001	1810000
CP000011_18	1800001	1910000
CP000011_19	1900001	2010000
CP000011_20	2000001	2110000
CP000011_21	2100001	2210000
CP000011_22	2200001	2310000
CP000011_23	2300001	2325379

Continuation (9 of 24) of CP000011 from base 800001 (CP000011 Burkholderia mallei ATCC 2

Query Match 13.3%; Score 133.4; DB 1; Length 110000;
Best Local Similarity 48.8%; Pred. No. 0.2;
Matches 488; Conservative 0; Mismatches 496; Indels 17; Gaps 4;

```
Qy          9 GCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGC 68
Db      30802 GCCCGCGCGCGCCTTCGCGATCCAGTCGGCGGTCTGCTCGATCAGCGCGGCGTCGAGCGC 30743

Qy         69 CAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCAC 128
Db      30742 GCGCCCCGCGACGTCGCGCACGAGCGCCTTGCATGCACGCACCGCCTGCGGGCCGTTCGC 30683

Qy        129 GAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTG 188
Db      30682 GCACAGCGCATCGGCGAGCTTCGCGACGCTCGCGTCGAGCGCATCG---GCCGGCACGCG 30626

Qy        189 CACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCC 248
Db      30625 CTCGTGCACGAAGCCGAGCTGCGCGGCCTTCGCGCAGTCGAACGCCTCGGCCGTCGCGAA 30566

Qy        249 GGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGG 308
Db      30565 ATAGCGGCGCGCCGCGCGCTCGCCCATCGCGCGCACGACGTACGGCGCGATCGTCGCCGG 30506

Qy        309 CAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAG 368
Db      30505 TATCAGCCCGAGGCGCGCCTCGGACAGGCAGAATTTCACGTCCTCGGCGGCAAGCGCGAT 30446

Qy        369 GCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAG 428
Db      30445 GTCGCACGCGGCGACGAGCCCGACGCCGCCCGCGTACGCGTCGCCGTGCACGCGCGCGAT 30386

Qy        429 CTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGAC 488
Db      30385 CACGGGCTTCGCGCAGCGATAGATCGCGTCGAGCATGCGCGCGAGCCGGCGCGCGTCGGC 30326

Qy        489 GAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGC 548
Db      30325 GCGGTTCTCGTCGTCGGAGAAGCCCGCCATCTTCTTCATCCAGTTCAGATCCGCGCCCGC 30266

Qy        549 GCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGC 608
Db      30265 GCAGAACGCGGGGCCGCGCGCG------GCGAGCACGACCGCGCGCAGCGACGGCTGCGC 30212

Qy        609 GAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAG 668
Db      30211 GTCGATGCGGTCGAACGCGGTCGTGAGCTCGGCGATCATCGTCTCGTTGAACGCGTTGCG 30152

Qy        669 CTCGGTGGGCC------GCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGT 722
Db      30151 CACGTCGGGGCGCGCGAGCGTGACGGTCGCGACGCGGCCGTCGTCGGTCACTTCGATCGT 30092

Qy        723 GACCTCGATCGCGGCCGCCTCGCGGCCGAGCA--GCCGCGCGAGCCTGTCCTCCAGCGCG 780
Db      30091 TTCGTAGCGCATCGTCGGCTCCTAGGCAGGCATCACATCCGGAACACGCCGAACCGCGTG 30032

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
Db      30031 TCCTCGATCGGCGCGTTTCGCGCGGCGGCAAGCGCGAGGCCCAGCACGTCGCGCGTCTGA 29972

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
Db      29971 GCGGGGTCGATCACGCCGTCGTCCCAGAGGCGCGCGCTCGCGTAATACGGATGGCCCTGG 29912

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCC 960
Db      29911 CGCTCGTACTGGTCGCGGATCGGCTGCTTGAACGCCTCTTCATCCTGCGCCGACCATTCG 29852

Qy        961 GCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001
Db      29851 CCGCCCTTCGCCTCGATCTCGTCGCGGCGCACGGTCGCGAG 29811
```

Search completed: April 28, 2006, 02:39:30
Job time : 5549.62 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 482.694 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_2000_3000
Perfect score: 1001
Sequence: 1 caggcgatgccggtgagcgg.....cccccgcgcaggagcgcgag 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1001	100.0	76994	12	ADP64454	Adp64454 Sorangium
c 2		242.2	24.2	2973	14	ACL73103	Acl73103 M. xanthu
3		242.2	24.2	9867	14	ACL64545	Acl64545 M. xanthu
c 4		174.2	17.4	4377	2	AAV23493	Aav23493 Pseudomon
c 5		174.2	17.4	4377	3	AAA13904	Aaa13904 Pseudomon
c 6		174.2	17.4	4377	4	AAF30869	Aaf30869 Pseudomon
c 7		174.2	17.4	4377	6	AAD22881	Aad22881 P. alcali
c 8		170.4	17.0	2526	11	ABD11744	Abd11744 Pseudomon
c 9		170.4	17.0	2991	11	ABD11843	Abd11843 Pseudomon
c 10		129	12.9	114955	2	AAX53491	Aax53491 Human ade
11		125.8	12.6	114955	2	AAX53491	Aax53491 Human ade
c 12		111.6	11.1	4770	8	AAL61172	Aal61172 Actinosyn
13		111.6	11.1	82746	8	AAL61224	Aal61224 Actinosyn
c 14		111	11.1	3513	8	ACA25950	Aca25950 Prokaryot
c 15		110.2	11.0	1029	2	AAV20098	Aav20098 Pseudomon
c 16		110.2	11.0	1029	3	AAA13890	Aaa13890 Pseudomon
c 17		110.2	11.0	1029	6	AAD22867	Aad22867 Pseudomon
c 18		109.6	10.9	3459	13	ADS55878	Ads55878 Bacterial
c 19		108.4	10.8	9975	8	AAL61173	Aal61173 Actinosyn
c 20		107.6	10.7	1172	6	ABQ31995	Abq31995 Oligonucl
c 21		107.6	10.7	1172	6	ABQ31996	Abq31996 Oligonucl
22		107.6	10.7	1172	6	ABQ31997	Abq31997 Oligonucl
23		107.6	10.7	1172	6	ABQ31994	Abq31994 Oligonucl
24		107	10.7	1172	6	ABQ31995	Abq31995 Oligonucl
25		107	10.7	1172	6	ABQ31996	Abq31996 Oligonucl
c 26		107	10.7	1172	6	ABQ31997	Abq31997 Oligonucl
c 27		107	10.7	1172	6	ABQ31994	Abq31994 Oligonucl
c 28		106.8	10.7	76994	12	ADP64454	Adp64454 Sorangium
c 29		105.8	10.6	6408	10	ADI23930	Adi23930 Streptomy
c 30		105.8	10.6	61944	10	ADI23920	Adi23920 Streptomy
31		105.6	10.5	138203	14	ADZ51725	Adz51725 FR-008 po
32		105.4	10.5	2217	11	ABD12116	Abd12116 Pseudomon
c 33		105.4	10.5	58857	3	AAA58471	Aaa58471 Nucleotid
c 34		104.4	10.4	14055	8	AAL61170	Aal61170 Actinosyn
c 35		104.4	10.4	88421	6	AAL40781	Aal40781 88421nt g
36		104	10.4	4542	10	ADC30607	Adc30607 Human nov
c 37		104	10.4	9222	8	AAL61171	Aal61171 Actinosyn
c 38		103.8	10.4	3354	8	AAL61181	Aal61181 Actinosyn
c 39		103.8	10.4	10002	8	ACA26601	Aca26601 Prokaryot
40		102.8	10.3	1164	8	AAL61217	Aal61217 Actinosyn
41		102.8	10.3	11905	8	AAL61225	Aal61225 Actinosyn
c 42		102.8	10.3	18876	10	ADI23928	Adi23928 Streptomy
c 43		102.6	10.2	37856	3	AAA11992	Aaa11992 S. cellul
c 44		102.6	10.2	132544	13	ADV99887	Adv99887 Nanchangm
c 45		102.4	10.2	6645	14	ACL72356	Acl72356 M. xanthu

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

PS Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC The present invention describes an isolated recombinant polynucleotide
CC (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC synthase (PKS) protein or a fragment comprising one or more domains of
CC the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC host cell (III) comprising (II), or (I) integrated into the cell
CC chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC the native disorazole PKS by the inactivation of one or more domain of
CC the disorazole PKS and/or addition of one or more domain of a non-
CC disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC expression system (VI) capable of producing a disorazole synthase domain
CC in a host cell, where the system comprises an encoding sequence for a
CC disorazole polyketide synthase domain, and the encoding sequence is
CC operably linked to control sequences effective in the cell to produce RNA
CC that is translated into the domain; (7) host cell modified to contain
CC (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC eliminate production of disorazole; (9) an isolated polypeptide encoded
CC by (I); (10) hybrid polyketide synthase comprising one or more
CC polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC of at least 200 base pairs with a sequence identical or substantially
CC identical to a protein encoding region of the 77294 base pair sequence of
CC SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC (I) is useful for producing disorazole polyketides. (III) is useful for
CC producing a polyketide, which involves growing (III) under conditions,
CC where a polyketide synthesised by a PKS comprising a protein encoded by
CC the polynucleotide molecule is produced in the cell. (I) is useful for
CC producing libraries of PKSs, where the polyketide producing colonies are
CC identified and isolated, and the produced polyketides are used
CC collectively in a panel to represent a library or assessed individually
CC for activity. The disorazole polyketide produced by (I) is useful in
CC molecular biology, chemistry, recombinant DNA technology, human and
CC veterinary medicine and agricultural applications. The disorazole
CC polyketide produced by (I) is formulated as a pharmaceutical composition,
CC useful for its activity against microorganisms, in a mammal. The present
CC sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

Query Match 100.0%; Score 1001; DB 12; Length 76994;
Best Local Similarity 100.0%; Pred. No. 1.1e-113;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2000 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 2059

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2060 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 2119

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2120 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 2179

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2180 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 2239

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2240 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 2299

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2300 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 2359

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2360 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 2419

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2420 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 2479

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2480 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 2539

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2540 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 2599

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2600 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 2659

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2660 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 2719

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2720 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 2779

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2780 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 2839

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2840 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 2899

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2900 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCC 2959

Qy        961 GCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001
              |||||||||||||||||||||||||||||||||||||||||
Db       2960 GCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 3000
```

RESULT 2
ACL73103/c
ID ACL73103 standard; DNA; 2973 BP.
XX
AC ACL73103;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus gene sequence, seq id 9566.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression;
KW gene; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.

```
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 9566; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 2973 BP; 357 A; 1039 C; 1095 G; 482 T; 0 U; 0 Other;

  Query Match             24.2%;  Score 242.2;  DB 14;  Length 2973;
  Best Local Similarity   57.0%;  Pred. No. 1.9e-21;
  Matches  561;  Conservative    0;  Mismatches  388;  Indels   36;  Gaps    5;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              | ||||||| | || ||||||||||| | ||| ||||||||||||||||| |  |||||
Db       2330 CTGGCGATGGCCGTCAGCGGGTTGCCAATCTCATGGGCCACGCCCGCGGCCACGCGCCCC 2271

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              |    |||||| |||||  | || || |   |||| || |  ||     | | |||
Db       2270 AGTGAGGCCAGCCGGTCCTGATGCGCCATGCGCGCGTCCACGGCCTTTCGTGCCGTGAAG 2211

Qy        121 TCCTCCACGAGCAGGACCACGC---CGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGC 177
              |||||||| || ||| ||| ||   ||  |||| | |  | |   ||    |||  ||||
Db       2210 TCCTCCACCAGGAGGGCCATGCCCTCGGACGAGCCCGTGCCCTGCTCCGCGGGAGAGAGC 2151

Qy        178 GCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCG 237
              |  |||||||  || || | |||| |    || ||||||   ||     |||||  |
Db       2150 CGCGAGCGGTGCAGCCGGAGAATGCGCTCGACGCCCGCCACGGTGACACGTGTCTCCGTG 2091

Qy        238 GCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCG 297
              | |                  ||     |  ||| |||    | ||  | ||||    |
Db       2090 TCTC------------------CGTCCACACCGCTGGCGAAACCGGACAGCAGCGGGCCC 2049

Qy        298 AACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACG 357
              | ||||| |||||    || || ||    ||    | |  |  | ||    |||||  |
Db       2048 CAGGGGGCGGGCAGGGCATCCAGGGGATGGCCACGCACGGTGTCCACCGGCACGCCGGAG 1989

Qy        358 AGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACC 417
              |||  ||||||   |  |||   || |  || | | || || |  |||||||||||||
Db       1988 AGCTGCTCGAGCGTCGCGTTCCAGATGACGACCTCACCATCCGGCCCCACGGCGCAGACG 1929

Qy        418 CCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTC 477
              || |  ||||| | | | || |     ||||||  | || |||| |    || ||| |   |
Db       1928 CCCACCGGGAGGTCCTCCAGGATGCGGCGCAGGTAGCGGCGCACGGCCTCCAGCGCGGGC 1869

Qy        478 GCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCC 537
              |||  |||    || | |  | || ||||    ||||||  ||| |   | ||||| |
Db       1868 ACCGGGCCTTGCAGTCCCCGCGCGTCGCG---AAGGCGCTCCTCGACGAAGCGGAGTTGT 1812

Qy        538 TCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCC 597
              ||||  || |||  |  |||||         ||  |  |   | |  | || ||||| ||
Db       1811 TCGGCCAGGGCGGTGCGCGCGC---------CGGGCTCCAGCCGCAGCGCCTCCTCCACC 1761

Qy        598 GCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGC 657
              |   | | |||||||||| ||||| | |||   |||||||| || | ||  |  ||  ||
Db       1760 GTTATGCGCGCGAGCACGGGGCCCAGCAGTCCCGACAGGTTTCGCTCCACGCCGTCGCGC 1701

Qy        658 AGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717
              ||   | | ||||||| ||||||   ||||||  |||  |    ||      || | ||
Db       1700 AGGCGGCGCAGCTCGGAGGGCCGGCGCTCGTCCTGCGGCAGCGACAGGGAATTCAGCGCC 1641

Qy        718 CGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGC 777
              ||    |||||   ||  || |||||     || |||||| | || ||| ||       | |
Db       1640 CGGTCCACCTCCGCCGTCGCGGCCTCCTCACCCAGCAGCGGGGCCAGCTTGCGGCGGAAC 1581

Qy        778 GCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTG 837
              ||    ||  |||  || |   | |||   || | | |||  ||  |||   ||| |||
Db       1580 TCG---TCGGGCGAGCCCGCCACCACGCCGCCCGCGGCGGGCGCGGCCTCCTCGCGTGTG 1524

Qy        838 CACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCG 897
              ||    || || |||||    |||||   || | |  |   |  |||||| |||     |
Db       1523 CAGGCCCGCGCGGCCTCTGCTTCCTCGGCGGACTGGCGCGTCGCCAGCGACACGCCGATG 1464

Qy        898 AGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGC 957
              |  | |   |||||||  | ||||||  |||| || | ||    ||| || ||  ||
Db       1463 AAGGCCAGCCCGTTGAGCGTGAGCGAGGCGAAGGTGGCGAAGCCCCAGGGCTCACTGTCG 1404

Qy        958 GCCGCGCCCGCCGGCGCGCCCGCGC 982
              |     ||   | ||||  ||  ||
Db       1403 GAGAACCCGAGCAGCGCCGCCATGC 1379


RESULT 3
ACL64545
ID   ACL64545 standard; DNA; 9867 BP.
XX
AC   ACL64545;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus DNA fragment, seq id 1008.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 1; SEQ ID NO 1008; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC   set of 1849 contig and singleton sequences comprising coding sequences,
CC   DNA replication elements, promoters and other regulatory elements from
CC   the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO
XX
SQ   Sequence 9867 BP; 1425 A; 3440 C; 3625 G; 1376 T; 0 U; 1 Other;

  Query Match             24.2%;  Score 242.2;  DB 14;  Length 9867;
```

```
  Best Local Similarity   57.0%;  Pred. No. 1.4e-21;
  Matches  561;  Conservative    0;  Mismatches  388;  Indels   36;  Gaps    5;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              | ||||||| | || |||||||||||| | ||| |||||||||||||||||| |  |||||
Db       1543 CTGGCGATGGCCGTCAGCGGGTTGCCAATCTCATGGGCCACGCCCGCGGCCACGCGCCCC 1602

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              |    |||||| |||||  | || || |   |||| || |  ||      |  | |||
Db       1603 AGTGAGGCCAGCCGGTCCTGATGCGCCATGCGCGCGTCCACGGCCTTTCGTGCCGTGAAG 1662

Qy        121 TCCTCCACGAGCAGGACCACGC---CGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGC 177
              |||||||| || ||| ||| ||   ||  |||| | |  | |   ||    |||  ||||
Db       1663 TCCTCCACCAGGAGGGCCATGCCCTCGGACGAGCCCGTGCCCTGCTCCGCGGGAGAGAGC 1722

Qy        178 GCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCG 237
              |  |||||||  || || | |||| |    || ||||||   ||     |||||  |
Db       1723 CGCGAGCGGTGCAGCCGGAGAATGCGCTCGACGCCCGCCACGGTGACACGTGTCTCCGTG 1782

Qy        238 GCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCG 297
              | |                    ||     |  ||| |||     | ||  | ||||    |
Db       1783 TCTC------------------CGTCCACACCGCTGGCGAAACCGGACAGCAGCGGGCCC 1824

Qy        298 AACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACG 357
              | |||||  |||||    || || ||    ||    | |  |  | ||    |||||  |
Db       1825 CAGGGGGCGGGCAGGGCATCCAGGGGATGGCCACGCACGGTGTCCACCGGCACGCCGGAG 1884

Qy        358 AGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACC 417
              |||  ||||||   |  |||   || |  || | | || || |  ||||||||||||||
Db       1885 AGCTGCTCGAGCGTCGCGTTCCAGATGACGACCTCACCATCCGGCCCCACGGCGCAGACG 1944

Qy        418 CCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTC 477
              || |  ||||| | | | || |    ||||||  | || |||| |   || ||| |   |
Db       1945 CCCACCGGGAGGTCCTCCAGGATGCGGCGCAGGTAGCGGCGCACGGCCTCCAGCGCGGGC 2004

Qy        478 GCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCC 537
              |||  |||    || | |  | || ||||    ||||||  |||  |   | |||||  |
Db       2005 ACCGGGCCTTGCAGTCCCCGCGCGTCGCG---AAGGCGCTCCTCGACGAAGCGGAGTTGT 2061

Qy        538 TCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCC 597
              ||||  || |||  |  |||||         ||  |  |   | |  | || ||||| ||
Db       2062 TCGGCCAGGGCGGTGCGCGCGC---------CGGGCTCCAGCCGCAGCGCCTCCTCCACC 2112

Qy        598 GCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGC 657
              |   | | ||||||||| ||||| | |||   |||||||||| || | ||  |   ||  ||
Db       2113 GTTATGCGCGCGAGCACGGGGCCCAGCAGTCCCGACAGGTTTCGCTCCACGCCGTCGCGC 2172

Qy        658 AGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717
              ||   | | ||||||| ||||||   ||||||  |||  |      ||      || | ||
Db       2173 AGGCGGCGCAGCTCGGAGGGCCGGCGCTCGTCCTGCGGCAGCGACAGGGAATTCAGCGCC 2232

Qy        718 CGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGC 777
              ||    |||||   ||  || |||||    || ||||||| | || ||| ||      | |
Db       2233 CGGTCCACCTCCGCCGTCGCGGCCTCCTCACCCAGCAGCGGGGCCAGCTTGCGGCGGAAC 2292

Qy        778 GCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTG 837
              ||    ||  |||   || |    | |||    || | | |||   ||  |||   ||| |||
Db       2293 TCG---TCGGGCGAGCCCGCCACCACGCCGCCCGCGGCGGGCGCGGCCTCCTCGCGTGTG 2349

Qy        838 CACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCG 897
              ||    || || |||||    |||||   || | |  |   |  ||||||| |||      |
Db       2350 CAGGCCCGCGCGGCCTCTGCTTCCTCGGCGGACTGGCGCGTCGCCAGCGACACGCCGATG 2409

Qy        898 AGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGC 957
              |  | |   |||||||  | ||||||  |||| || | ||    ||| || ||  ||
Db       2410 AAGGCCAGCCCGTTGAGCGTGAGCGAGGCGAAGGTGGCGAAGCCCCAGGGCTCACTGTCG 2469

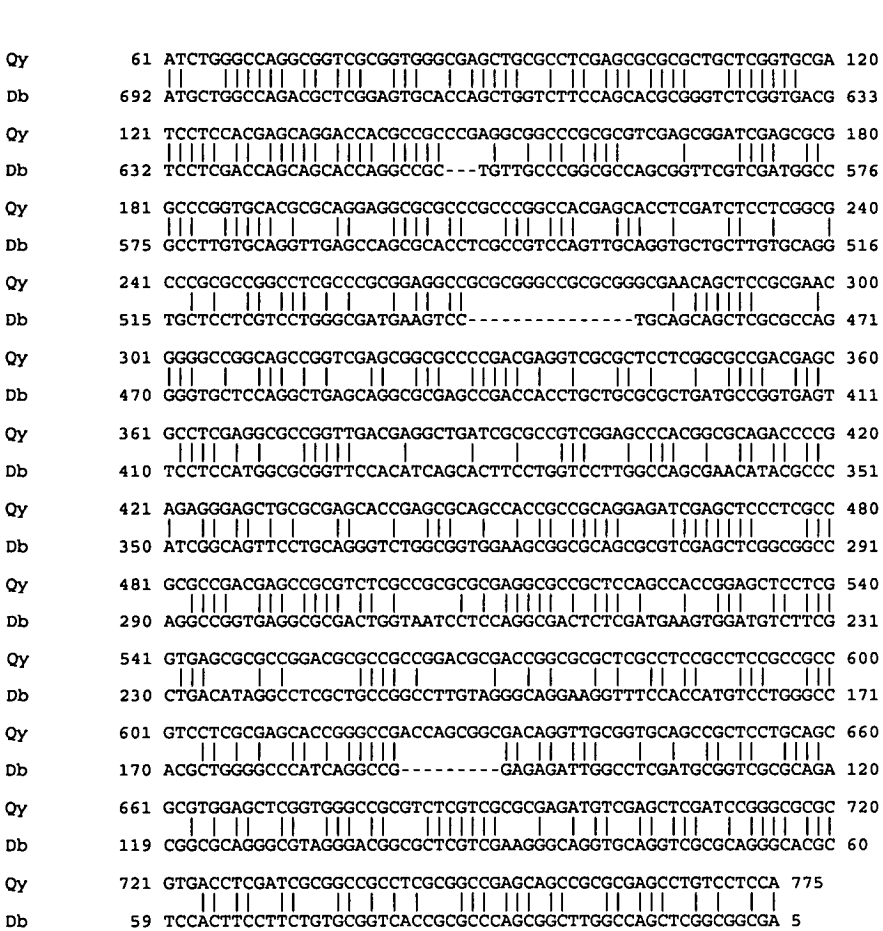
Qy        958 GCCGCGCCCGCCGGCGCGCCCGCGC 982
              |     ||   | ||||  ||  ||
Db       2470 GAGAACCCGAGCAGCGCCGCCATGC 2494
```

RESULT 4

```
AAV23493/c
ID   AAV23493 standard; DNA; 4377 BP.
XX
AC   AAV23493;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Pseudomonas LipQ, LipR, OrfV operon.
XX
KW   Kinase; LipQ; LipR; lipase expression regulator; DNA binding regulator;
KW   sigma 54 promoter; secretion factor; lux-box binding element;
KW   orfV-box binding element; regulation cascade; ss.
XX
OS   Pseudomonas alcaligenes.
XX
FH   Key             Location/Qualifiers
FT   CDS             348. .1377
FT                   /*tag=  a
FT                   /product= "LipQ"
FT   CDS             1443. .2874
FT                   /*tag=  b
FT                   /product= "LipR"
FT   CDS             3531. .4377
FT                   /*tag=  c
FT                   /product= "OrfZ"
FT                   /note= "no stop codon given"
XX
PN   WO9806836-A2.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US014450.
XX
PR   16-AUG-1996;   96US-00699092.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 1998-159528/14.
DR   P-PSDB; AAW53824, AAW53825, AAW53840.
XX
PT   Nucleic acids encoding proteins involved in the lipase regulation cascade
PT   from P. alcaligenes - useful for controlling production and secretion of
PT   heterologous proteins in P. alcaligenes.
XX
PS   Disclosure; Page 42-43; 106pp; English.
XX
CC   This sequence encodes the OrfZ secretion factor, LipQ and lipR kinases of
CC   Pseudomonas alcaligenes. The DNA represents a nucleic acid of the
CC   invention. The nucleic acids encode: (a) a kinase from a Pseudomonad that
CC   regulates the expression of a lipase; (b) a DNA binding regulator from a
CC   Pseudomonad that regulates the expression of a lipase; (c) a Pseudomonas
CC   alcaligenes upstream activating sequence; (d) a P. alcaligenes sigma 54
CC   promoter that regulates the expression of a lipase; (e) a P. alcaligenes
CC   secretion factor selected from XcpP, Q, R, S, T, U, V, W, X, Y and Z, and
CC   OrfV, X, Y; (f) a P. alcaligenes lux-box binding element; and (g) a orfV-
CC   box binding element. The nucleic acids represent parts of a regulation
CC   cascade, comprising at its heart a kinase and a DNA binding regulator.
CC   These sequences can be used for the production of heterologous proteins
CC   in a host cell. The cascade also comprises secretion factors which can
CC   enhance the secretion of produced proteins
XX
SQ   Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;

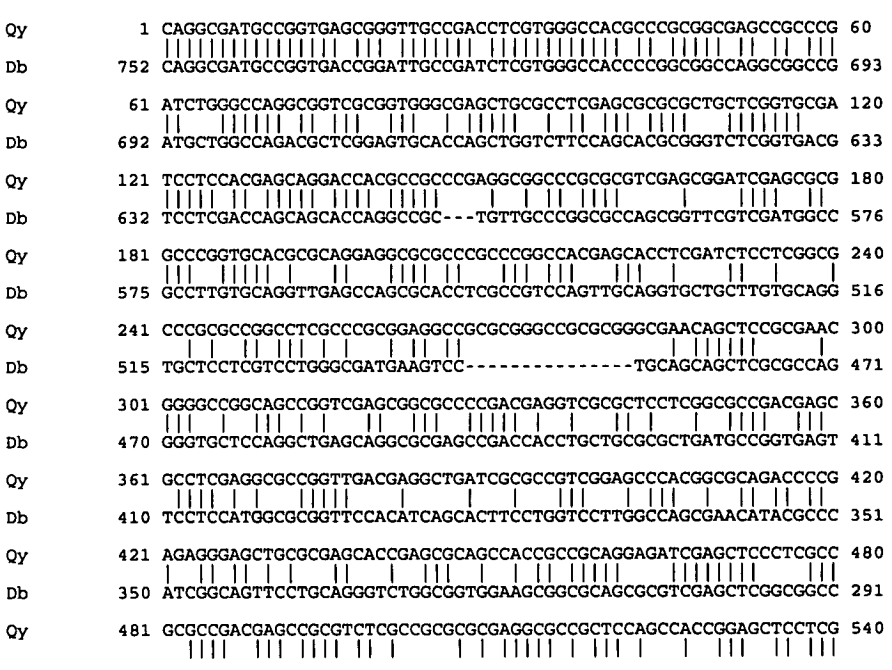
  Query Match            17.4%;  Score 174.2;  DB 2;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 3e-13;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              |||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693
```



RESULT 5
AAA13904/c
ID AAA13904 standard; DNA; 4377 BP.
XX
AC AAA13904;
XX
DT 01-AUG-2000 (first entry)
XX
DE Pseudomonas alcaligenes nucleotide sequence SEQ ID NO:28.
XX
KW Pseudomonas alcaligenes; expression; lipase regulation cascade; kinase;
KW DNA binding regulator; polymerase; promoter; secretion factor; XcpP;
KW XcpQ; XcpR; XcpS; XcpT; XcpU; XcpV; XcpW; XcpX; XcpY; XcpZ; OrfV; OrfX;
KW OrfY; OrfZ; LipQ; LipR; upstream activating sequence; detergent;
KW cleaning formulation; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN US6048710-A.
XX
PD 11-APR-2000.
XX
PF 15-AUG-1997; 97US-00911853.
XX
PR 16-AUG-1996; 96US-00699092.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Gerritse G, Quax WJ;
XX
DR WPI; 2000-316896/27.
XX
PT Expression vector for producing heterologous proteins in host cells
PT comprises a nucleic acid encoding a kinase and a DNA binding regulator
PT which hybridizes under stringent conditions to nucleic acid.
XX
PS Example 3; Fig 4; 133pp; English.
XX
CC The present invention describes an expression vector comprising a nucleic
CC acid encoding a kinase and a DNA binding regulator which hybridises under
CC stringent conditions to a nucleic acid isolated from Pseudomonas
CC alcaligenes. Also described are: (1) an isolated plasmid comprising the
CC above expression vector; (2) a method of transforming a host cell
CC comprising adding the above plasmid to host cells under appropriate
CC conditions; (3) a transformed host cell comprising the above expression
CC vector; and (4) a method for producing a protein comprising the steps of
CC obtaining a host cell comprising the above expression vector and further
CC comprising nucleic acid encoding the protein, and culturing the host cell
CC under conditions for the expression of protein. The expression vector of
CC the present invention can be used for producing heterologous proteins in
CC host cells, particularly, lipase in Pseudomonas. Lipases produced can be
CC used in detergents and cleaning formulations in industrial processes. The
CC invention provides a higher production level and efficiently express a
CC heterologous protein. The present sequence represents a Pseudomonas
CC alcaligenes nucleotide sequence which contains the LipQ, LipR and OrfZ
CC sequences, from the present invention
XX
SQ Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;

Query Match 17.4%; Score 174.2; DB 3; Length 4377;
Best Local Similarity 54.8%; Pred. No. 3e-13;
Matches 425; Conservative 0; Mismatches 323; Indels 27; Gaps 3;



```
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600

Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660

Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720

Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775

Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 6
AAF30869/c
ID   AAF30869 standard; DNA; 4377 BP.
XX
AC   AAF30869;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Pseudomonas alcaligenes LipQ, LipR and OrfZ gene region.
XX
KW   LipQ; kinase; LipQ; DNA binding regulator; OrfZ; polymerase; lipase; ds.
XX
OS   Pseudomonas alcaligenes.
XX
FH   Key             Location/Qualifiers
FT   protein_bind    257. .275
FT                   /*tag= a
FT                   /standard_name= "lux-box"
FT   CDS             349. .1377
FT                   /*tag= b
FT                   /gene= "lipQ"
FT                   /product= "kinase"
FT                   /note= "encodes AAB82253; this coding region is
FT                   specifically claimed in Claim 1"
FT   CDS             1443. .2858
FT                   /*tag= c
FT                   /gene= "lipR"
FT                   /product= "DNA binding regulator"
FT                   /note= "encodes AAB2254"
FT   stem_loop       2930. .2966
FT                   /*tag= d
FT                   /function= "terminator"
FT   protein_bind    3055. .3067
FT                   /*tag= e
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3059. .3070
FT                   /*tag= f
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3085. .3097
FT                   /*tag= g
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3115. .3147
FT                   /*tag= h
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3118. .3130
FT                   /*tag= i
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3210. .3213
FT                   /*tag= j
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3327. .3339
FT                   /*tag= k
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3349. .3361
FT                   /*tag= l
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   protein_bind    3352. .3364
FT                   /*tag= m
FT                   /standard_name= "orfV-box"
FT                   /bound_moiety= "regulator OrfV"
FT   CDS             3531. .4376
FT                   /*tag= n
FT                   /partial
FT                   /gene= "orfZ"
FT                   /product= "polymerase"
FT                   /note= "encodes AAB82255"
XX
PN   US6225106-B1.
XX
PD   01-MAY-2001.
XX
PF   06-JAN-2000; 2000US-00479409.
XX
PR   16-AUG-1996;   96US-00699092.
PR   15-AUG-1997;   97US-00911853.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 2001-315684/33.
DR   P-PSDB; AAB82253, AAB82254, AAB82255.
XX
PT   Novel isolated nucleic acid encoding kinase from Pseudomonas that can
PT   regulate expression of lipase, useful in expression systems for
PT   production of lipase which is useful in detergents and cleaning
PT   formulations.
XX
PS   Claim 1; Fig 4A-G; 131pp; English.
XX
CC   The present sequence is derived from overlapping sequences of cosmids 71,
CC   201, 505 and 726 that include open reading frames of lipQ (claimed), lipR
CC   and orfZ from Pseudomonas alcaligenes. The cosmids were isolated
CC   following construction of a cosmid library from P. alcaligenes DNA in
CC   Escherichia coli, transformation of the cosmid library into P.
CC   alcaligenes overexpressing lipase, and selection of cosmids stimulating
CC   lipase expression. The lipQ gene, identified by homology to known
CC   sequences, encodes a kinase (see AAB72253) that regulates lipase
CC   expression. The lipR gene encodes a DNA binding regulator (see AAB82254).
CC   The incomplete orfZ gene encodes a polymerase. Another isolated cosmid
CC   (see AAF30870) includes genes encoding P. alcaligenes secretion factors.
CC   A new expression system is provided which comprises components of a
CC   lipase regulation cascade including the kinase, DNA binding regulator,
CC   polymerase, a promoter, an upstream activating sequence, and the
CC   secretion factors. Plasmids and transformed cells are provided, and also
CC   host cells which further comprise a nucleic acid encoding a desired
CC   protein, especially an esterase, hydrolase, lipase, isomerase, mutase,
CC   transferase, kinase or phosphatase (claimed). A hyperproducing strain can
CC   be developed
XX
SQ   Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 174.2;  DB 4;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 3e-13;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60

Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693
```

```
Qy          61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
               ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db         692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy         121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
               ||||| || ||||| |||| |||||    |  | || ||||     |    ||||  ||
Db         632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy         181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
               |||  ||||| |   ||   |||| ||   ||| |||   |||  |    ||  |    |
Db         575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy         241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                 |  ||  ||| | |   | || ||                  | ||||||     |
Db         515 TGCTCCTCGTCCTGGGCGATGAAGTCC---------------TGCAGCAGCTCGCGCCAG 471

Qy         301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
               |||  |  |||  | |   ||  |||  ||||| |  |   || |   |  ||||  |||
Db         470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

Qy         361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
                ||||  |  |   |||||     |     |   |   |||   | |||  |  || || ||
Db         410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy         421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
               |  ||  ||  | |   ||   |  |||  |  | || |||||    ||||||||    |||
Db         350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy         481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
                ||||  |||  |||| || |     | | ||||| | |||  |   |  |||  || |||
Db         290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy         541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
                |||    | |      |||| |       |  | |   | |  || ||   |||   |||
Db         230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy         601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
               || | |  || | ||||       || || |||   |  |  || ||  ||||
Db         170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy         661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
               | | ||  ||  ||| ||   |||||||   |  | ||  || |||  | |||| |||
Db         119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy         721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
                || ||  ||   || | | |   ||| ||| ||   || |||  | |  | |
Db          59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 7
AAD22881/c
ID   AAD22881 standard; DNA; 4377 BP.
XX
AC   AAD22881;
XX
DT   26-FEB-2002  (first entry)
XX
DE   P. alcaligenes DNA comprising cosmids #71, #201, #505 and #726.
XX
KW   DNA binding regulator; lipase regulation cascade; kinase; polymerase;
KW   promoter; upstream activating factor; secretion factor; detergent;
KW   cleaning formulation; industrial process; LipQ; LipR; OrfZ; ds.
XX
OS   Pseudomonas alcaligenes.
XX
FH   Key             Location/Qualifiers
FT   misc_binding    257. .275
FT                   /*tag=  a
FT                   /bound_moiety= "Lux-box (auto-inducer)"
FT   CDS             349. .1377
FT                   /*tag=  b
FT                   /product= "LipQ protein"
FT   CDS             1443. .2858
FT                   /*tag=  c
FT                   /product= "LipR protein"
FT   terminator      2930. .2942
FT                   /*tag=  d
FT   terminator      2954. .2966
FT                   /*tag=  e
FT   misc_binding    3055. .3067
FT                   /*tag=  f
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3058. .3070
FT                   /*tag=  g
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3085. .3097
FT                   /*tag=  h
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3115. .3127
FT                   /*tag=  i
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3118. .3130
FT                   /*tag=  j
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3200. .3212
FT                   /*tag=  k
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3327. .3339
FT                   /*tag=  l
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3349. .3361
FT                   /*tag=  m
FT                   /bound_moiety= "OrfV regulator"
FT   misc_binding    3351. .3364
FT                   /*tag=  n
FT                   /bound_moiety= "OrfV regulator"
FT   CDS             3531. .4376
FT                   /*tag=  o
FT                   /product= "OrfZ protein"
FT                   /note= "CDS does not include stop codon"
FT                   /partial
XX
PN   US6313283-B1.
XX
PD   06-NOV-2001.
XX
PF   06-JAN-2000; 2000US-00479453.
XX
PR   16-AUG-1996;   96US-00699092.
PR   15-AUG-1997;   97US-00911853.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 2002-024912/03.
DR   P-PSDB; AAE13654, AAE13655, AAE13670.
XX
PT   Polynucleotide encoding DNA binding regulator, useful for regulating
PT   expression of lipase especially in pseudomonad, and lipase useful in
PT   detergents and other cleaning formulations and in various industrial
PT   processes.
XX
PS   Example 3; Fig 4; 132pp; English.
XX
CC   The present invention relates to a polynucleotide encoding a DNA binding
CC   regulator that can regulate the expression of a lipase. The invention
CC   also relates to an expression system comprising components of Pseudomonas
CC   alcaligenes lipase regulation cascade which includes kinases, DNA binding
CC   regulators, polymerases, promoters, upstream activating factors and
CC   secretion factors. DNA binding regulators of the invention are useful for
CC   regulating the expression of a lipase, where the lipase is useful in
CC   detergents and other cleaning formulations as well as a number of
CC   industrial processes. Various components of lipase regulation cascade is
CC   useful in expression methods and systems designed for the production of
```

```
CC   heterologous proteins. The present sequence is P. alcaligenes DNA
CC   comprising the cosmids #71, #201, #505 and #726 that includes the open
CC   reading frames of LipQ, LipR and a part of OrfZ proteins
XX
SQ   Sequence 4377 BP; 876 A; 1539 C; 1317 G; 645 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 174.2;  DB 6;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 3e-13;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||    |||||||
Db        692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              |||||  || |||||  |||| |||||     |   |  ||  ||||      |     ||||   ||
Db        632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  |||||  |    ||   |||| ||   |||  |||   |||   |     ||   |     |
Db        575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                |  |  || ||| | |    | || ||                      | ||||||      |
Db        515 TGCTCCTCGTCCTGGGCGATGAAGTCC---------------TGCAGCAGCTCGCGCCAG 471

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||  |  ||| | |    ||  |||  |||||  |   |    || |    |  ||||  |||
Db        470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||  | |    |||||     |      |    |   |||    | |||   |  || || ||
Db        410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || | |    ||    |  |||   |  | || |||||     ||||||||     |||
Db        350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  ||| |||| || |      | | |||||  | ||| |   |  |||  || |||
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              |||    | |       |||| |      |  | |    | |  || ||    |||   |||
Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              || | |  || | |||||          || || |||    |  |  || ||  ||||
Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | | ||  ||  ||| ||   |||||||    |  | ||  || |||  | |||| |||
Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
                || ||  ||   || | |  ||| ||| ||   || |||  | |  | |
Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 8
ABD11744/c
ID   ABD11744 standard; DNA; 2526 BP.
XX
AC   ABD11744;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #10348.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO78173.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 10348; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 2526 BP; 400 A; 920 C; 755 G; 451 T; 0 U; 0 Other;

  Query Match             17.0%;  Score 170.4;  DB 11;  Length 2526;
  Best Local Similarity   51.5%;  Pred. No. 1e-12;
  Matches  486;  Conservative    0;  Mismatches  431;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db       2313 CAGGCGATGCCGGTGATCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCTAGGCGGCCG 2254

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||    |||||||| |||  |||| | ||||    | || | |    |    ||||| |
Db       2253 ATGGAAGCCAGGCGCTCGGAGTGGACCAGCTTGTCTTCCAACAGTTGGGTTTCGGTGAGG 2194

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              || ||||| || || |||| |||||     |  | || ||||    | |    ||||  ||
Db       2193 TCTTCCACCAGTAGCACCAGGCCGC---TGTTGCCCGGCGCCAGGGGTTCCTCGATGGCC 2137

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  |||||  |  |||   ||| |   | ||| |   || ||        ||  |   |
Db       2136 GCCTTGTGCAGGTTCAGCCAGCGGGTGTGGCCGTCGTAGACCAGGCGCTGCTTGTGCAGG 2077

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                | || |||   || |  | ||                    ||  | | |||  ||    |
Db       2076 TGCTCGTCGGGGGCGTCGATGAA---------------ACGTTCCAGCAGGCCCTTCCAG 2032

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||  | ||||          ||  |||  |||||||  |   ||  |  || | |||  |||
Db       2031 GGTTCGGGCAAGGCCGACAGGCGCGAACCGACGACCTTCTGCGCGCCGACCCCGGTGAGT 1972
```

```
Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||||  |   |||||     |     |   |   |||    | ||| || ||||||  ||
Db       1971 TCCTCGATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGGGAGCAGACGCCC 1912

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || |   ||   |   ||  |  | || ||||||   ||||| ||    |||
Db       1911 ATCGGCAGGTCTTGCAGGGTCTGCCGGTGGTAGCGGCGCAGGGTGTCGAGTTCGGCAGCC 1852

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              |||    ||||| |  |     |  | | ||||| | ||| |   |  |||  ||||||
Db       1851 AGGCCAGTCAGCCGGGAGTGATAGTCCTCCAGGCGGCTCTCGATGAAGTGGATGTCCTCG 1792

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||    ||      ||| |         |    | || || |    ||
Db       1791 GTGACATAGCTTTCTTCGCTGGTCTTGTAGGGAAGGAAGGTCTCCACGATGTC------- 1739

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              |  ||||    || || | ||||   || |||||    |  |  || ||  |||
Db       1738 --CTGGGCGACACTGGGTCCCATCAGCCCGGAAAGGTTCGCCTCGATGCGGTCGCGCAAG 1681

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | ||  |   ||| ||    ||||||    |  | || ||| |||  | ||||  ||
Db       1680 CGACGCAGGGCATAGGGGCGACGTTCGTCGAAGGGCAGGTGGAGATCGCGCAGGGCCTGC 1621

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||  ||   || | ||  |  ||| ||| ||   || | |  |    | | | |
Db       1620 TCGACCTCCTTCTGCGCGGTCTTGGCGCCCAGCGGCTTGGCCAACTGGCTGGCGAACTCC 1561

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              |        ||  |    ||||      ||   |||  |  | |     || |||
Db       1560 TGCGGCGAACCGGCGAACAGTTCGCGGCGTTGCGGACGGCGTACGTTGTCCACCGCGCAG 1501

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              |||  ||  |    |||||   |   | |  |   | |||||| || |    ||
Db       1500 GCCTCGGCGGCGCCTTTTTCCTCGTCGCTGGCCTCGGTGAACAGCGACACCAGGGTGAAT 1441

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
              |    ||||| | ||  |||||    | ||  ||  ||   |||
Db       1440 ACCAGCACGTTGGCCGCCAGCGAGGACAGCGCCGCCAGGTGCCA 1397
```

RESULT 9
ABD11843/c
ID ABD11843 standard; DNA; 2991 BP.
XX
AC ABD11843;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #10447.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO78272.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 10447; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 2991 BP; 489 A; 1072 C; 900 G; 530 T; 0 U; 0 Other;

```
  Query Match             17.0%;  Score 170.4;  DB 11;  Length 2991;
  Best Local Similarity   51.5%;  Pred. No. 9.5e-13;
  Matches  486;  Conservative    0;  Mismatches  431;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              |||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db       2369 CAGGCGATGCCGGTGATCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCTAGGCGGCCG 2310

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||    |||||||| |||  |||| | ||||   | || | |    |    |||||| |
Db       2309 ATGGAAGCCAGGCGCTCGGAGTGGACCAGCTTGTCTTCCAACAGTTGGGTTTCGGTGAGG 2250

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              || ||||| || || |||| |||||     |  | || ||||   | |    ||||  ||
Db       2249 TCTTCCACCAGTAGCACCAGGCCGC---TGTTGCCCGGCGCCAGGGGTTCCTCGATGGCC 2193

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |  |||   ||| |   | ||| |   || ||        ||  |    |
Db       2192 GCCTTGTGCAGGTTCAGCCAGCGGGTGTGGCCGTCGTAGACCAGGCGCTGCTTGTGCAGG 2133

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              |  || |||   || |   | |                    ||  | | |||  ||    |
Db       2132 TGCTCGTCGGGGGCGTCGATGAA--------------ACGTTCCAGCAGGCCCTTCCAG 2088

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||  | ||||         ||  |||  |||||||  |   || |  ||  | |||  |||
Db       2087 GGTTCGGGCAAGGCCGACAGGCGCGAACCGACGACCTTCTGCGCGCCGACCCCGGTGAGT 2028

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||||  |   |||||     |     |   |||    | ||| || ||||||  ||
Db       2027 TCCTCGATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGGGAGCAGACGCCC 1968

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || |      ||   |   ||  |  | || ||||||    ||||| ||    |||
Db       1967 ATCGGCAGGTCTTGCAGGGTCTGCCGGTGGTAGCGGCGCAGGGTGTCGAGTTCGGCAGCC 1908

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              |||    ||||| |  |     |  | | ||||| | ||| |   |  |||  ||||||
Db       1907 AGGCCAGTCAGCCGGGAGTGATAGTCCTCCAGGCGGCTCTCGATGAAGTGGATGTCCTCG 1848

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||    ||      ||| |         |    | || || |    ||
Db       1847 GTGACATAGCTTTCTTCGCTGGTCTTGTAGGGAAGGAAGGTCTCCACGATGTC------- 1795

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              |  ||||    || || | ||||   || |||||    |  |  || ||  |||
Db       1794 --CTGGGCGACACTGGGTCCCATCAGCCCGGAAAGGTTCGCCTCGATGCGGTCGCGCAAG 1737
```

```
Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | ||| |   ||| ||    ||||||   | | || ||| |||  | ||||  ||
Db       1736 CGACGCAGGGCATAGGGGCGACGTTCGTCGAAGGGCAGGTGGAGATCGCGCAGGGCCTGC 1677

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||  ||   || | || |  ||| ||| ||   || | |   |    | | | |
Db       1676 TCGACCTCCTTCTGCGCGGTCTTGGCGCCCAGCGGCTTGGCCAACTGGCTGGCGAACTCC 1617

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              |          || |      ||||      ||   |||  |  | |     || |||
Db       1616 TGCGGCGAACCGGCGAACAGTTCGCGGCGTTGCGGACGGCGTACGTTGTCCACCGCGCAG 1557

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              |||  ||  |    |||||  |   | |  |   | ||||| || |    ||
Db       1556 GCCTCGGCGGCGCCTTTTTCCTCGTCGCTGGCCTCGGTGAACAGCGACACCAGGGTGAAT 1497

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
              |    ||||| | || |||||    | || ||  ||   |||
Db       1496 ACCAGCACGTTGGCCGCCAGCGAGGACAGCGCCGCCAGGTGCCA 1453
```

```
RESULT 10
AAX53491/c
ID   AAX53491 standard; DNA; 114955 BP.
XX
AC   AAX53491;
XX
DT   05-JUL-1999  (first entry)
XX
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
XX
OS   Synthetic.
XX
PN   WO9913886-A1.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US019419.
XX
PR   17-SEP-1997;   97US-0059160P.
PR   09-JUN-1998;   98US-00093972.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 1999-229400/19.
XX
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction.
XX
PS   Disclosure; Page 37; 120pp; English.
XX
CC   The specification describes antisense oligonucleotides (AAX52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene initiation
CC   codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'
CC   -end and the juxta-section between coding and non-coding regions and all
CC   segments of RNAs encoding proteins associated with one or more diseases,
CC   conditions or mixtures. The antisense oligonucleotides may be derived
CC   from sequences AAX55272-74. These multiple target oligonucleotides
CC   (specifically AAX55180-271) can be used for the antisense treatment of
CC   diseases and conditions. Typical diseases and conditions are those
CC   associated with impaired respiration and inflammation, including lung
CC   diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,
CC   acute asthma, allergies, asthma, impeded respiration, respiratory
CC   distress syndrome, pain, cystic fibrosis, pulmonary hypertension,
CC   pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary
CC   disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
CC   colon cancer, breast cancer, lung cancer, pancreatic cancer,
CC   hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
CC   well as all types of cancers which may metastasize or have metastasized
CC   to the lungs, including breast and prostate cancer
XX
SQ   Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

  Query Match             12.9%;  Score 129;  DB 2;  Length 114955;
  Best Local Similarity   36.0%;  Pred. No. 3.8e-08;
  Matches  357;  Conservative  105;  Mismatches  523;  Indels    7;  Gaps    4;

Qy         10 CCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCC 69
              ||||   ||||   |||   :    ||||  ||| ||||||  ||| :         ||||
Db     105217 CCGGCCGGCGGCGCGCCCNNHNNNSCGGCCCGGCCGGCGGCGCGCCCVNNHNNNSCGGCC 105158

Qy         70 AGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACG 129
              ||| | ||  | |  |:  :  :|     ||  || || ||   :|  :   ||
Db     105157 CGGCCGGCGGCGCGCCCVGNNHNNNSCGGCCCGGCCGGCGGCGCGCCCVGGNNHNNNSCG 105098

Qy        130 AGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGC 189
              | || |  || ||| |  :|    :   :|| |  |      || ||||| ||:||
Db     105097 GCCCGGCCGGCGGCGCGCCCVGGCNNHNNNSCGGCCCGGCCGGCG-GCGCGCCCVGGCCN 105039

Qy        190 ACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCG 249
              :  : ||    ||   || |  ||| :|  |: :    |  | ||||   ||   ||
Db     105038 NHNNNSCGGCCCGGCCGGCGGCGCCCVGGCCVNNHNNNSCGGCCCGGCCGGCGGCGCG 104979

Qy        250 GCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGC 309
              || |||:|      :||  | |||||   | |||  |:|  |:|| :   :    |
Db     104978 CCCVGGCCVGNNHNNNSCGGCCCGGCCGGCGGCGCGCCCVGGCCVGCNNHNNNSCGGCCC 104919

Qy        310 AGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGG 369
              |||||  | ||| | :  | :     :   :  ||   |   :   ::|  |: ||
Db     104918 GGCCGGCGGCGCGCCCVGGCCVGCGNNHNNNSGGCCVGCGGNNHNNNSVGGCCVGCGGNN 104859

Qy        370 CGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGC 429
              :   :  | |:  |   :   : |:| | :      :   :|  :  ||:  |  :
Db     104858 HNNNSCVGGCCVGCGGNNHNNNSCCVGGCCVGCGGNNHNNNSCCCVGGCCVGCGGNNHNN 104799

Qy        430 TGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACG 489
              :  | :|  |:| |   :    |  ||:   :|      :     ||: |||: ||
Db     104798 NSGCCCVGGCCVGCGGNNHNNNSCGCCCVGGCCVGCGGNNHNNNSGCGCCCVGGCCVGCG 104739

Qy        490 AGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCG 549
                :    | |||| :|  |:  |   :   :  ||  ||:|  |  |    :   :|
Db     104738 GNNHNNNSCGCGCCCVGGCCVGCGGNNHNNNSGCGCGCCCVGGCCVGCGGNNHNNNSGGC 104679

Qy        550 CCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGC- 608
              ||  |: |||: |||       :|||||||| |:   |:||        :    | |||
Db     104678 GCGCCCVGGCCVGCGGNNHNNNSCGGCGCGCCCVGGCCVGCGGNNHNNNSGCGGCGCGCC 104619

Qy        609 --GAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGG 666
                |  |: |||  :   : ||||||        || :      :   :| || |||||
Db     104618 CVGGCCVGCGGNNHNNNSGGCGGCGCGCCCVGGCCVGCGGNNHNNNSCGGCGGCGCGCCC 104559

Qy        667 AGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACC 726
              :|  |:| ||  :    |  | |||||     | | :||    :   :|| |||   |
Db     104558 VGGCCVGCGGNNHNNNSCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNSGCCGGCGGCGCG 104499

Qy        727 TCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACG 786
              | :  :||| |  :   :| |  || ||     | :| | :|    :   : | | |
Db     104498 CCCVGGCCVGCGGNNHNNNSGGCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNSCGGCCGG 104439

Qy        787 CTCGACGCCACGGTCGC--GCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGC 844
              |    |||| :|| |:   |  :   |||  ||| |  | | |:|  |:
Db     104438 CGGCGCGCCCVGGCCVGCGGNNHNNNSCCGGCCGGCGGCGCGCCCVGGCCVGCGGNNHNN 104379
```

```
Qy        845 GGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCG 904
              : ||| |  |||   | ||| ||||:||||      :  |  | ||    || ||  :|
Db     104378 NSCCCGGCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNSGCCCGGCCGGCGGCGCGCCCVG 104319

Qy        905 CGCCGTTG-ACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCG 963
              |:|  |  :   : ||    |  || || |  |:  | :| |   :   :|| | ||
Db     104318 GCCVGCGGNNHNNNSGGCCCGGCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNSCGGCCCG 104259

Qy        964 CCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAG 995
              ||| ||||||||||: |||: ||   :   :
Db     104258 GCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNS 104227


RESULT 11
AAX53491
ID   AAX53491 standard; DNA; 114955 BP.
XX
AC   AAX53491;
XX
DT   05-JUL-1999  (first entry)
XX
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
XX
OS   Synthetic.
XX
PN   WO9913886-A1.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US019419.
XX
PR   17-SEP-1997;   97US-0059160P.
PR   09-JUN-1998;   98US-00093972.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 1999-229400/19.
XX
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction.
XX
PS   Disclosure; Page 37; 120pp; English.
XX
CC   The specification describes antisense oligonucleotides (AAX52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene initiation
CC   codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'
CC   -end and the juxta-section between coding and non-coding regions and all
CC   segments of RNAs encoding proteins associated with one or more diseases,
CC   conditions or mixtures. The antisense oligonucleotides may be derived
CC   from sequences AAX55272-74. These multiple target oligonucleotides
CC   (specifically AAX55180-271) can be used for the antisense treatment of
CC   diseases and conditions. Typical diseases and conditions are those
CC   associated with impaired respiration and inflammation, including lung
CC   diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,
CC   acute asthma, allergies, asthma, impeded respiration, respiratory
CC   distress syndrome, pain, cystic fibrosis, pulmonary hypertension,
CC   pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary
CC   disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
CC   colon cancer, breast cancer, lung cancer, pancreatic cancer,
CC   hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
CC   well as all types of cancers which may metastasize or have metastasized
CC   to the lungs, including breast and prostate cancer
XX
SQ   Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

  Query Match             12.6%;  Score 125.8;  DB 2;  Length 114955;
  Best Local Similarity   34.5%;  Pred. No. 9.2e-08;
  Matches  334;  Conservative  101;  Mismatches  527;  Indels    5;  Gaps    2;

Qy         25 CCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGG 84
              | | |:   :   |  :| ||: |||| |||||| |    |:   :  | | :|  . :||
Db     104336 CGGGCSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCCGGGSNNNDNNCCGCBGGCCBGG 104395

Qy         85 GCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCG 144
              ||| || ||    || :   :  | ||: |  : |   | || |  || |:   :  |||
Db     104396 GCGCGCCGCCGGCCGGSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCCGSNNNDNNCCG 104455

Qy        145 CCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGC 204
              |: |  : ||| |||     |  :   :      :||||:||       ||| || :
Db     104456 CBGGCCBGGGCGCGCCGCCGGCCSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCSNNND 104515

Qy        205 GCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAG 264
              ||||: | |:  | || || |   :   :  |||: || : |||    | |  :
Db     104516 NNCCGCBGGCCBGGGCGCGCCGCCGGSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGSNNN 104575

Qy        265 GCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGC 324
              :  |||:||||: |  |     | :          |:|| |:||  ||    |:   :
Db     104576 DNNCCGCBGGCCBGGGCGCGCCGCCSNNNDNNCCGCBGGCCBGGGCGCGCCGCSNNNDNN 104635

Qy        325 GCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGG 384
              | |:| |  || ||||| :   :  ||| :|  |:   || | |:   :       ||
Db     104636 CCGCBGGCCBGGGCGCGCCGSNNNDNNCCGCBGGCCBGGGCGCGCCSNNNDNNCCGCBGG 104695

Qy        385 CTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAG 444
              | :       |:   :  ||  :||| : | | | :   :      : ||  |  |
Db     104696 CCBGGGCGCGCSNNNDNNCCGCBGGCCBGGGCGCGSNNNDNNCCGCBGGCCBGGGCGCSN 104755

Qy        445 CGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCG 504
                 || |:| | : || |     :   | |: |||: | |:   :  |   :  | : |
Db     104756 NNDNNCCGCBGGCCBGGGCGSNNNDNNCCGCBGGCCBGGGCSNNNDNNCCGCBGGCCBGG 104815

Qy        505 CGCGCGAGGCGCCGCTCCAGCC---ACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCG 561
              :  :   |||:|  |: |:     | | : ||: :   :   |||:|| |::   :
Db     104816 GSNNNDNNCCGCBGGCCBGGSNNNDNNCCGCBGGCCBGSNNNDNNCCGCBGGCCBSNNND 104875

Qy        562 CCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCG 621
                |: | |:   :  | |:   |:|    |||||||| ||  ||  :      ||:|
Db     104876 NNCCGCBGGCCSNNNDNNCGCBGGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNGCBG 104935

Qy        622 ACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGC 681
              || | |    | |   || | || |:   :    | |  |: || || | |  |||||
Db     104936 GCCBGGGCGCGCCGCCGGCCGGGCCGSNNND--NNCBGGCCBGGGCGCGCCGCCGGCCGG 104993

Qy        682 GTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGCC 741
              | | :   :  :|   :|     ||     | ||  :        |  :| |  ||||
Db     104994 GCCGSNNNDNNBGGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNGGCCBGGGCGCGCC 105053

Qy        742 TCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTC 801
              |  | | |  | :       |  :  |  | ||  ||       |          | :|| |
Db     105054 GCCGGCCGGGCCGSNNNDNNGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNCCBGGGC 105113

Qy        802 GCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCC 861
              ||||   |  ||| |  |:   :  |: |||||     | | |||||||:   :  :
Db     105114 GCGCCGCCGGCCGGGCCGSNNNDNNCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNBGGG 105173

Qy        862 TCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGC 921
              |   |||||| ||  | :       |     |||   | || ||||       | ||
Db     105174 CGCGCCGCCGGCCGGGCCGSNNNDNNGGGCGCGCCGCCGGCCGGGCCGSNNNDNNGGCGC 105233

Qy        922 GACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCG 981
              | | |   |   |   :       |  |    |||   ||  :      ||||||  ||
```

```
Db     105234 GCCGCCGGCCGGGCCGSNNNDNNGCGCGCCGCCGGCCGGGCCGSNNNDNNCGCGCCGCCG 105293

Qy        982 CCCCCGC 988
              ||  ||
Db     105294 GCCGGGC 105300
```

RESULT 12
AAL61172/c
ID AAL61172 standard; DNA; 4770 BP.
XX
AC AAL61172;
XX
DT 22-SEP-2003 (first entry)
XX
DE Actinosynnema pretiosum polyketide synthase (PKS) gene #3.
XX
KW Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;
KW gene; ds.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
DR WPI; 2003-493374/46.
XX
PT Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS Disclosure; Page 59-61; 160pp; English.
XX
CC The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumour agent. The present sequence is A. pretiosum
CC ansamitocin gene cluster I polyketide synthase (PKS) gene
XX
SQ Sequence 4770 BP; 485 A; 1947 C; 1809 G; 529 T; 0 U; 0 Other;

Query Match 11.1%; Score 111.6; DB 8; Length 4770;
Best Local Similarity 47.4%; Pred. No. 1.1e-05;
Matches 463; Conservative 0; Mismatches 504; Indels 9; Gaps 4;

```
Qy           3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
               || | |||     |     ||  | || | ||| |   | ||  | | || ||   | |
Db        4092 GGTGGTGCTCTCGTCGCGGGCGACGGCGTCGAGCAGCGCGGACACCGCCAGGTCCGGGGG 4033

Qy          63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATC 122
               |       | || |    |||||||||||| ||  |||     |  | |   || ||
Db        4032 CATCAGGCGCAGGCCAGCCTGGGCGAGCTGGGCGGCGA---TGTCCGGGGTGGCGCTGGA 3976

Qy         123 CTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGC 182
               |    ||     | ||  | ||||| ||||| |    | | |||   |  || |||||
Db        3975 CATGCCGCCGCCGTCCCAGGCGCCCCAGGCGACGGAGGTGGCGACGCGGCCGTGCGCGCG 3916

Qy         183 CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCC 242
               ||| ||  |  |||   |      ||  ||    ||    | | |     || | ||||
Db        3915 GCGGCGCTGGGCCAGTGCGTCGAGGCGGGCGTTGGCGGCGGCGTAGGCGCACTGGCCGCC 3856

Qy         243 CGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGG 302
               |||   | |||||||| ||  | |       ||   |||   |||
Db        3855 GCTGCCCCACACGCCCGCGTTGGAGGAGAACAGGACGAACGCGTCGAGCGGGCGCCCGCC 3796

```

```
Qy         303 GGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGC 362
               | ||      |  | ||   || ||||     | |    ||| | |||   || || |
Db        3795 CAGCAGCTCGTCGAGGTGCTCGGCGCCGAGCGCCTTGGCGGCCTGGACGCGCGCGTGCTC 3736

Qy         363 CTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAG 422
               |||| ||     || | ||| |  | ||  | ||| || |||| | |||   ||  ||
Db        3735 CTCGGGGGTGGTGTCGGCGACGGGGGTCTGGGCGTTGGCGCCCGCCGCGTGCACGACGGC 3676

Qy         423 AGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGC 482
               | || |||| ||  | |     ||| |   | |||||    |||| | |       |
Db        3675 GGTGACCTGCTCGCCCGCGCGCTCCAGGCCGGTCAGCAGGGCGGCGAGCGCGGCGCGGTC 3616

Qy         483 GCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGT 542
               | |||||   | |   || ||  | || |||     |||||  |   || ||  |||
Db        3615 GGCGACGTCGCAGG---CGGCGGTGGTGACGCGGGCGCCCAGCGCCTCCAGGTCGGCGGC 3559

Qy         543 GAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGT 602
               |||||||   |   ||  | | || | | |||  |||   |     | |   |  | ||
Db        3558 GAGCGCGGGCGCGCCGGGGGCCGA-GGGGCCGCGGCGGCTGGTGAGGACGAGGTGGTCGG 3500

Qy         603 CCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGC 662
               |  || | ||  || | || |  ||  ||  |          |  | | |||  | | |
Db        3499 CGCCGAGGGCGGCGAGGCGGCGGGCCACGTGCGCGGCGACGCCGCCGGTTCCGCCGGTGA 3440

Qy         663 GTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGT 722
               |  |   | |  || |   |  | |||   |          | || | |||| ||
Db        3439 TCAGGACCGTGCCGCGCGGGGTCCACGGGGCGGCCGGGGCCTGCGGTTCGGCGCGCACGA 3380

Qy         723 GACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGC 782
               | |  ||  ||    |||| ||| ||    ||  |  |  |||  |||||  ||  |
Db        3379 GCCTGCGCGCGAACACGCCCTGCGCGCGGATGGCGACCTGGTCCTCACCTCCGTCGTCGA 3320

Qy         783 CACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAG 842
               | |      || |  || ||   ||    ||| || || |   |   ||| |  | |
Db        3319 CGCCGGGTCCGGCGAGGACGGCGGCGAGCCCGCGGCGCTGGCGCGGGTCGGCGGGGTCGG 3260

Qy         843 GCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGT 902
               || || |  || | || |   |||||  ||| ||  ||||| |||||  | |  |
Db        3259 CGAGGTCGACCAGCCCGCCCCAGCGGCCG--GGGTGCTCCAGCGCGACGACGCGGCCCAG 3202

Qy         903 CGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGC 962
               | | | || ||||  | |     |  || ||||          ||||  ||| |   | ||
Db        3201 TCCCCAGGTGGCGGCCTGGGCGGGGTCCGGCGGGCCGTCGTGCGGGGAGATCGCGACGGC 3142

Qy         963 GCCCGCCGGCGCGCCC 978
               || |  |  |||| |
Db        3141 CCCGGTGGTGGCGCAC 3126
```

RESULT 13
AAL61224
ID AAL61224 standard; DNA; 82746 BP.
XX
AC AAL61224;
XX
DT 22-SEP-2003 (first entry)
XX
DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.
XX
KW Maytansinoid; ansamitocin; antitumour; gene; ds.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW) UNIV WASHINGTON.

XX
PI Floss HG, Yu T, Leistner E;
XX
DR WPI; 2003-493374/46.
XX
PT Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS Claim 7; Page 105-152; 160pp; English.
XX
CC The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumour agent. The present sequence is A. pretiosum
CC ansamitocin ansamitocin biosynthetic gene cluster I
XX
SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;

Query Match 11.1%; Score 111.6; DB 8; Length 82746;
Best Local Similarity 47.4%; Pred. No. 5.2e-06;
Matches 463; Conservative 0; Mismatches 504; Indels 9; Gaps 4;

```
Qy          3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
              || | |||     |    ||  | || | ||| |   | ||  | | || ||   | |
Db      67888 GGTGGTGCTCTCGTCGCGGGCGACGGCGTCGAGCAGCGCGGACACCGCCAGGTCCGGGGG 67947

Qy         63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATC 122
              |       |  || |    ||||||||||| ||  |||     |  | |    || ||
Db      67948 CATCAGGCGCAGGCCAGCCTGGGCGAGCTGGGCGGCGA---TGTCCGGGGTGGCGCTGGA 68004

Qy        123 CTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGC 182
              |    ||      | ||  | ||||| ||||| |     | | |||    |  || |||||
Db      68005 CATGCCGCCGCCGTCCCAGGCGCCCCAGGCGACGGAGGTGGCGACGCGGCCGTCGCGCGG 68064

Qy        183 CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCC 242
              ||| ||  |  |||   |      ||  ||    ||    | | |     || | ||||
Db      68065 GCGGCGCTGGGCCAGTGCGTCGAGGCGGGCGTTGGCGGCGGCGTAGGCGCACTGGCCGCC 68124

Qy        243 CGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGG 302
              |||   | ||||||||  ||  | |       ||   |||   |||
Db      68125 GCTGCCCCACACGCCCGCGTTGGAGGAGAACAGGACGAACGCGTCGAGCGGGCGCCCGCC 68184

Qy        303 GGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGC 362
              | ||     |  | ||   || ||||     | |     ||| | |||   || || |
Db      68185 CAGCAGCTCGTCGAGGTGCTCGGCGCCGAGCGCCTTGGCGGCCTGGACGCGCGCGTGCTC 68244

Qy        363 CTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAG 422
              |||| ||     || | ||| |  | ||  | ||| || |||| | |||   ||  ||
Db      68245 CTCGGGGGTGGTGTCGGCGACGGGGGTCTGGGCGTTGGCGCCCGCCGCGTGCACGACGGC 68304

Qy        423 AGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGC 482
              | || |||| ||  | |       ||| |     | |||||    |||||  |        |
Db      68305 GGTGACCTGCTCGCCCGCGCGCTCCAGGCCGGTCAGCAGGGCGGCGAGCGCGGCGCGGTC 68364

Qy        483 GCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGT 542
              | |||||   |  |   || ||   | || |||      |||||  |   || ||  |||
Db      68365 GGCGACGTCGCAGG---CGGCGGTGGTGACGCGGGCGCCCAGCGCCTCCAGGTCGGCGGC 68421

Qy        543 GAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGT 602
              |||||||   |   ||  | | || | | |||  |||   |     | |   |  | ||
Db      68422 GAGCGCGGGCGCGCCGGGGGCCGA-GGGGCCGCGGCGGCTGGTGAGGACGAGGTGGTCGG 68480

Qy        603 CCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGC 662
              |  || | ||  || | || |  ||  ||     |   | | |||  | | |
Db      68481 CGCCGAGGGCGGCGAGGCGGCGGGCCACGTGCGCGGCGACGCCGCCGGTTCCGCCGGTGA 68540

Qy        663 GTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGT 722
              |  |   |  | || |    |  |||     |          | || | |||| ||
Db      68541 TCAGGACCGTGCCGCGCGGGGTCCACGGGGCGGCCGGGGCCTGCGGTTCGGCGCGCACGA 68600

Qy        723 GACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGC 782
              | |  ||  ||    ||||  |||  ||    ||  |  |  |||  |||||  ||  |
```

```
Db      68601 GCCTGCGCGCGAACACGCCCTGCGCGCGGATGGCGACCTGGTCCTCACCTCCGTCGTCGA 68660

Qy        783 CACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAG 842
              | |       || |  || ||   ||     ||| || || |      |    ||| |  | |
Db      68661 CGCCGGGTCCGGCGAGGACGGCGGCGAGCCCGCGGCGCTGGCGCGGGTCGGCGGGGTCGG 68720

Qy        843 GCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGT 902
              || || |  || | || |   ||||  ||| ||  |||| |||||  | |  |
Db      68721 CGAGGTCGACCAGCCCGCCCCAGCGGCCG--GGGTGCTCCAGCGCGACGACGCGGCCCAG 68778

Qy        903 CGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGC 962
              | | | || ||||  | |     |  || ||||        ||||  ||| |   | ||
Db      68779 TCCCCAGGTGGCGGCCTGGGCGGGGTCCGGCGGGCCGTCGTGCGGGGAGATCGCGACGGC 68838

Qy        963 GCCCGCCGGCGCGCCC 978
              || |  |  |||| |
Db      68839 CCCGGTGGTGGCGCAC 68854
```

RESULT 14
ACA25950/c
ID ACA25950 standard; DNA; 3513 BP.
XX
AC ACA25950;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #7607.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Burkholderia mallei.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU22080.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 13820; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)

```
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3513 BP; 635 A; 1171 C; 1286 G; 421 T; 0 U; 0 Other;

  Query Match             11.1%;  Score 111;  DB 8;  Length 3513;
  Best Local Similarity   48.1%;  Pred. No. 1.4e-05;
  Matches  476;  Conservative    0;  Mismatches  495;  Indels   18;  Gaps    5;

Qy         26 CGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGG 85
              || |||||| | || || ||| |||||| || ||         |  ||           | |
Db       2815 CGGCCTCGTCGACCTCGGCCGTCGCGAGCTGCTCGGCGAACTGCTCGCCCGTCATACGCG 2756

Qy         86 CGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGC 145
              |   |||| |||  ||     |||    | |||||||   ||   || | | |  ||
Db       2755 CCGCCTGCTCCTTCAGTTGCAGCTCGGTGATGCGATCGCGCAGCGGCTGCAGCGAGCGCT 2696

Qy        146 CCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCG 205
              | | ||||  |||||| |||   |   || ||     |      |       |||   ||
Db       2695 CGGCGGCGAGCCGCGCCTCGTCGGCCGCGCGCAGCTTCGCCGTGAGATCGTCGAGCTCCG 2636

Qy        206 CCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGG 265
              | ||| |||| ||||| |||    || | | |||| | |  ||  | || ||     |
Db       2635 CGCGCGCGGCGCCGAGCGCCTGCTCCTTCGCCGCGCGCACTTCGAGCGCGTCCTGCAGGC 2576

Qy        266 CCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCG 325
              | | | | || |   |  ||   | |  | ||| |   ||  ||      ||||||| ||
Db       2575 CGGTGTGCGCGGTCTGCTCGTTGATCGTCTCGAGCTCCGCGCGCGCATCCTCGAGCGACG 2516

Qy        326 CCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGC 385
              || ||||| | |||  || ||||    ||||  |  |  | |   | | | | ||| ||
Db       2515 CCGCGACGCGCTCGGCCTGCTCGTGCGCGACCTGGATCGAGCGCTTCAGCTCGTCGATGC 2456

Qy        386 TGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGC 445
              | |||||    |    || | ||||| ||  | ||   | |  |           |  ||
Db       2455 GGTTCGCGGACT---GGCGCGCGGCGAAGCGCGCGTCGGTGGCCGCGCGCTCGCGCTCGC 2399

Qy        446 GCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGC 505
              ||  |  | | |||  |      |||  |    |         ||| |     ||    |
Db       2398 GCGCCTCCTGGCGCGCGTTCGTCAGCGTCTCGTCGAGCGATTCGAACGCGAGCTGATTGT 2339

Qy        506 GCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTG-AGCGCGCCGGACGCGCCGCCG 564
              | ||| ||||  || ||       ||||||  ||  |  |||| |   |  |||| |||
Db       2338 CCTCGAAGCGCGCCTGCAATTCGGCGAGCTCGGCGTCGTGGCGCTCGAAATTCGCCTCCG 2279

Qy        565 GACGCGACCGGCGCGCTCG-----CCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGC 619
                 ||  |||| ||     ||||  || |||  |  ||  | ||||  ||  ||
Db       2278 ACTCCGCGCGCAGCGCACGCTGCTCCTCGATCTGCGCGCCGATCTCCTCGAGTTCCTCGC 2219

Qy        620 CGACCAGCG------GCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGG 673
              || | |||      ||  |   | ||| |   ||  || |   |||  | | |  ||
Db       2218 GGATCTGCGTGCTGCGCTGCGTATAGCGCTCGTGCGCCTGCGTGAGCTTGAGCACGTCCA 2159

Qy        674 TGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCG 733
              |  || |||  |   |||||    |  |  |||||  | | ||||||    |    | ||
Db       2158 TCTGCAGCGCGTGCACGCGCTGCGTCGCACGCTCGGCCTGCGCGCGCACTTCGGTCAGCG 2099
```

```
Qy        734 CGGCCGCCTCGCGGCCGAGC---AGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCG 790
              |   || | |  |   | ||      ||||  |  |  || ||  |   ||| ||   |
Db       2098 CCTGCGACGCCTGCGTGTGCGCGGCTTCCGCGCGGATCGCCGCCGCCTTCGCCTCGTCGG 2039

Qy        791 ACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCG 850
              ||   |  | |         ||||  | |    |  |   |        |  ||| ||
Db       2038 ACAGGAGCGCCTGCGCGCGCACCTGGCGCGTCAGGTTCTCGATTTCCTGCGCGCGCGCGA 1979

Qy        851 CCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGT 910
              |  || | ||| |  ||      |         |||   |||  |  | ||||  || |||
Db       1978 GCATGCCCGCCTGCTCGGAGTCGGCCGCATAGAGCTGCACGCCCGAGCGCGTGACGACGT 1919

Qy        911 TGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCG 970
              | | ||   |   |||||||    |  |  |  |  |  ||| |||| |||  ||||
Db       1918 GGCCCGCCTTCACGACGAACGCGCCGCCCTGCGGCAATTGCATGCGCGCGGCGAGCGCCT 1859

Qy        971 GCGCGCCCGCGCCCCCGCGCAGGAGCGCG 999
              |||||    || |  ||    |  | || |
Db       1858 GCGCGAGGTCGTCGGCGACGAACACCGTG 1830
```

```
RESULT 15
AAV20098/c
ID   AAV20098 standard; DNA; 1029 BP.
XX
AC   AAV20098;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Pseudomonas LipQ kinase coding sequence.
XX
KW   Kinase; LipQ; LipR; lipase expression regulator; DNA binding regulator;
KW   sigma 54 promoter; secretion factor; lux-box binding element;
KW   orfV-box binding element; regulation cascade; ss.
XX
OS   Pseudomonas alcaligenes.
XX
PN   WO9806836-A2.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US014450.
XX
PR   16-AUG-1996;   96US-00699092.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Gerritse G,  Quax WJ;
XX
DR   WPI; 1998-159528/14.
DR   P-PSDB; AAW53824.
XX
PT   Nucleic acids encoding proteins involved in the lipase regulation cascade
PT   from P. alcaligenes - useful for controlling production and secretion of
PT   heterologous proteins in P. alcaligenes.
XX
PS   Claim 3; Page 21-22; 106pp; English.
XX
CC   This sequence encodes the LipQ kinase of Pseudomonas alcaligenes. The DNA
CC   represents a nucleic acid of the invention. The nucleic acids encode: (a)
CC   a kinase from a Pseudomonad that regulates the expression of a lipase;
CC   (b) a DNA binding regulator from a Pseudomonad that regulates the
CC   expression of a lipase; (c) a Pseudomonas alcaligenes upstream activating
CC   sequence; (d) a P. alcaligenes sigma 54 promoter that regulates the
CC   expression of a lipase; (e) a P. alcaligenes secretion factor selected
CC   from XcpP, Q, R, S, T, U, V, W, X, Y and Z, and OrfV, X, Y; (f) a P.
CC   alcaligenes lux-box binding element; and (g) a orfV-box binding element.
CC   The nucleic acids represent parts of a regulation cascade, comprising at
CC   its heart a kinase and a DNA binding regulator. These sequences can be
CC   used for the production of heterologous proteins in a host cell. The
CC   cascade also comprises secretion factors which can enhance the secretion
CC   of produced proteins
XX
```

```
SQ   Sequence 1029 BP; 192 A; 352 C; 341 G; 144 T; 0 U; 0 Other;

  Query Match            11.0%;  Score 110.2;  DB 2;  Length 1029;
  Best Local Similarity  59.9%;  Pred. No. 2.4e-05;
  Matches  203;  Conservative    0;  Mismatches  133;  Indels    3;  Gaps    1;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              |||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        404 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 345

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        344 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 285

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||| || ||||| |||| |||||     |  | || ||||      |    ||||  ||
Db        284 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 228

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |   ||   |||| ||   ||| |||   |||  |    ||  |    |
Db        227 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 168

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              | |  || ||| | |   | || ||     |   | ||||   |    |||||  |
Db        167 TGCTCCTCGTCCTGGGCGATGAAGTCCTGCAGCAGCTCGCGCCAGGGGTGCTCCAGGCTG 108

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCG 339
              |    ||   | | | | | ||  | ||  || | ||
Db        107 AGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCG 69


Search completed: April 27, 2006, 07:59:55
Job time : 487.694 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 206.521 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_2000_3000
Perfect score: 1001
Sequence: 1 caggcgatgccggtgagcgg.....cccccgcgcaggagcgcgag 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	242.2	24.2	2973	3	US-09-902-540-9566	Sequence 9566, Ap
	2	242.2	24.2	9867	3	US-09-902-540-1008	Sequence 1008, Ap
c	3	174.2	17.4	4377	3	US-08-911-853-28	Sequence 28, Appl
c	4	174.2	17.4	4377	3	US-09-479-409-28	Sequence 28, Appl
c	5	174.2	17.4	4377	3	US-09-479-453-28	Sequence 28, Appl
c	6	170.4	17.0	2526	3	US-09-252-991A-10348	Sequence 10348, A
c	7	170.4	17.0	2991	3	US-09-252-991A-10447	Sequence 10447, A
c	8	110.2	11.0	1029	3	US-08-911-853-1	Sequence 1, Appli
c	9	110.2	11.0	1029	3	US-09-479-409-1	Sequence 1, Appli
c	10	110.2	11.0	1029	3	US-09-479-453-1	Sequence 1, Appli
	11	105.4	10.5	2217	3	US-09-252-991A-10720	Sequence 10720, A
c	12	105.4	10.5	58857	3	US-09-477-962-1	Sequence 1, Appli
c	13	102.4	10.2	6645	3	US-09-902-540-8819	Sequence 8819, Ap
c	14	102.4	10.2	10210	3	US-09-902-540-938	Sequence 938, App
c	15	101.8	10.2	1839	3	US-09-758-759-150	Sequence 150, App
c	16	101.8	10.2	109519	3	US-09-758-759-1	Sequence 1, Appli
c	17	101.4	10.1	1779	3	US-09-902-540-3312	Sequence 3312, Ap
	18	101.4	10.1	8563	3	US-09-902-540-3318	Sequence 3318, Ap
c	19	101.4	10.1	15351	3	US-09-902-540-1154	Sequence 1154, Ap
	20	101.2	10.1	1683	3	US-09-724-797-21	Sequence 21, Appl
c	21	101.2	10.1	77536	3	US-09-410-551B-1	Sequence 1, Appli
c	22	101.2	10.1	77536	3	US-09-940-316B-1	Sequence 1, Appli
c	23	101	10.1	2109	3	US-09-902-540-8825	Sequence 8825, Ap
c	24	101	10.1	8765	3	US-09-902-540-939	Sequence 939, App
c	25	99.2	9.9	4257	2	US-08-690-473-1	Sequence 1, Appli
c	26	99.2	9.9	4257	3	US-09-259-821A-1	Sequence 1, Appli
c	27	99.2	9.9	4257	3	US-08-843-659-1	Sequence 1, Appli
c	28	99.2	9.9	4257	3	US-09-825-288A-1	Sequence 1, Appli
c	29	99	9.9	47981	3	US-09-679-279-1	Sequence 1, Appli
c	30	98.6	9.9	1320	3	US-09-724-797-85	Sequence 85, Appl
c	31	98.4	9.8	15872	3	US-09-105-537-1	Sequence 1, Appli
c	32	98.4	9.8	15872	3	US-09-091-609-1	Sequence 1, Appli
c	33	98.4	9.8	15872	3	US-09-091-609-3	Sequence 3, Appli
c	34	97.4	9.7	71989	3	US-09-443-501A-2	Sequence 2, Appli
c	35	97.4	9.7	71989	3	US-09-724-889A-2	Sequence 2, Appli
c	36	97.4	9.7	71989	3	US-09-724-876-2	Sequence 2, Appli
c	37	97.2	9.7	1683	3	US-09-724-797-21	Sequence 21, Appl
c	38	97.2	9.7	3957	3	US-10-237-551-193	Sequence 193, App
	39	97.2	9.7	154746	3	US-09-827-688-8	Sequence 8, Appli
c	40	97.2	9.7	154746	3	US-09-827-688-8	Sequence 8, Appli
c	41	96.4	9.6	1431	3	US-09-724-797-87	Sequence 87, Appl
c	42	95.6	9.6	3033	3	US-09-724-797-81	Sequence 81, Appl
c	43	95.6	9.6	20235	2	US-07-642-734C-3	Sequence 3, Appli
c	44	95.6	9.6	20235	3	US-08-439-009A-3	Sequence 3, Appli
	45	95.4	9.5	109519	3	US-09-758-759-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-902-540-9566/c
; Sequence 9566, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9566
; LENGTH: 2973
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9566

Query Match 24.2%; Score 242.2; DB 3; Length 2973;
Best Local Similarity 57.0%; Pred. No. 3.8e-29;
Matches 561; Conservative 0; Mismatches 388; Indels 36; Gaps 5;

```
Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              | ||||||| | || |||||||||||| | ||| |||||||||||||||||| |  |||||
Db       2330 CTGGCGATGGCCGTCAGCGGGTTGCCAATCTCATGGGCCACGCCCGCGGCCACGCGCCCC 2271

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              |    |||||| |||||  | || || |   |||| || |  ||     |  | |||
Db       2270 AGTGAGGCCAGCCGGTCCTGATGCGCCATGCGCGCGTCCACGGCCTTTCGTGCCGTGAAG 2211

Qy        121 TCCTCCACGAGCAGGACCACGC---CGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGC 177
              ||||||| || ||| ||| ||    ||  |||| | |  | |  |    ||     |||  ||||
Db       2210 TCCTCCACCAGGAGGGCCATGCCCTCGGACGAGCCCGTGCCCTGCTCCGCGGGAGAGAGC 2151

Qy        178 GCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCG 237
                |  |||||||  || || | |||| |    || ||||||   ||     |||||  |
Db       2150 CGCGAGCGGTGCAGCCGGAGAATGCGCTCGACGCCCGCCACGGTGACACGTGTCTCCGTG 2091

Qy        238 GCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCG 297
                | |                  ||    | ||| |||    | ||  | ||||    |
Db       2090 TCTC-----------------CGTCCACACCGCTGGCGAAACCGGACAGCAGCGGGCCC 2049

Qy        298 AACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACG 357
```

```
Db       2048 CAGGGGGCGGGCAGGGCATCCAGGGGATGGCCACGCACGGTGTCCACCGGCACGCCGGAG 1989

Qy        358 AGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACC 417

Db       1988 AGCTGCTCGAGCGTCGCGTTCCAGATGACGACCTCACCATCCGGCCCCACGGCGCAGACG 1929

Qy        418 CCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTC 477

Db       1928 CCCACCGGGAGGTCCTCCAGGATGCGGCGCAGGTAGCGGCGCACGGCCTCCAGCGCGGGC 1869

Qy        478 GCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCC 537

Db       1868 ACCGGGCCTTGCAGTCCCCGCGCGTCGCG---AAGGCGCTCCTCGACGAAGCGGAGTTGT 1812

Qy        538 TCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCC 597

Db       1811 TCGGCCAGGGCGGTGCGCGCGC---------CGGGCTCCAGCCGCAGCGCCTCCTCCACC 1761

Qy        598 GCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGC 657

Db       1760 GTTATGCGCGCGAGCACGGGGCCCAGCAGTCCCGACAGGTTTCGCTCCACGCCGTCGCGC 1701

Qy        658 AGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717

Db       1700 AGGCGGCGCAGCTCGGAGGGCCGGCGCTCGTCCTGCGGCAGCGACAGGGAATTCAGCGCC 1641

Qy        718 CGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGC 777

Db       1640 CGGTCCACCTCCGCCGTCGCGGCCTCCTCACCCAGCAGCGGGGCCAGCTTGCGGCGGAAC 1581

Qy        778 GCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTG 837

Db       1580 TCG---TCGGGCGAGCCCGCCACCACGCCGCCCGCGGCGGGCGCGGCCTCCTCGCGTGTG 1524

Qy        838 CACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCG 897

Db       1523 CAGGCCCGCGCGGCCTCTGCTTCCTCGGCGGACTGGCGCGTCGCCAGCGACACGCCGATG 1464

Qy        898 AGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGC 957

Db       1463 AAGGCCAGCCCGTTGAGCGTGAGCGAGGCGAAGGTGGCGAAGCCCCAGGGCTCACTGTCG 1404

Qy        958 GCCGCGCCCGCCGGCGCGCCCGCGC 982

Db       1403 GAGAACCCGAGCAGCGCCGCCATGC 1379
```

RESULT 2
US-09-902-540-1008
; Sequence 1008, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1008
; LENGTH: 9867
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(9867)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1008

Query Match 24.2%; Score 242.2; DB 3; Length 9867;
Best Local Similarity 57.0%; Pred. No. 3.5e-29;
Matches 561; Conservative 0; Mismatches 388; Indels 36; Gaps 5;

```
Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60

Db       1543 CTGGCGATGGCCGTCAGCGGGTTGCCAATCTCATGGGCCACGCCCGCGGCCACGCGCCCC 1602

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120

Db       1603 AGTGAGGCCAGCCGGTCCTGATGCGCCATGCGCGCGTCCACGGCCTTTCGTGCCGTGAAG 1662

Qy        121 TCCTCCACGAGCAGGACCACGC---CGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGC 177

Db       1663 TCCTCCACCAGGAGGGCCATGCCCTCGGACGAGCCCGTGCCCTGCTCCGCGGGAGAGAGC 1722

Qy        178 GCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCG 237

Db       1723 CGCGAGCGGTGCAGCCGGAGAATGCGCTCGACGCCCGCCACGGTGACACGTGTCTCCGTG 1782

Qy        238 GCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCG 297

Db       1783 TCTC------------------CGTCCACACCGCTGGCGAAACCGGACAGCAGCGGGCCC 1824

Qy        298 AACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACG 357

Db       1825 CAGGGGGCGGGCAGGGCATCCAGGGGATGGCCACGCACGGTGTCCACCGGCACGCCGGAG 1884

Qy        358 AGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACC 417

Db       1885 AGCTGCTCGAGCGTCGCGTTCCAGATGACGACCTCACCATCCGGCCCCACGGCGCAGACG 1944

Qy        418 CCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTC 477

Db       1945 CCCACCGGGAGGTCCTCCAGGATGCGGCGCAGGTAGCGGCGCACGGCCTCCAGCGCGGGC 2004

Qy        478 GCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCC 537

Db       2005 ACCGGGCCTTGCAGTCCCCGCGCGTCGCG---AAGGCGCTCCTCGACGAAGCGGAGTTGT 2061

Qy        538 TCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCC 597

Db       2062 TCGGCCAGGGCGGTGCGCGCGC---------CGGGCTCCAGCCGCAGCGCCTCCTCCACC 2112

Qy        598 GCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGC 657

Db       2113 GTTATGCGCGCGAGCACGGGGCCCAGCAGTCCCGACAGGTTTCGCTCCACGCCGTCGCGC 2172

Qy        658 AGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717

Db       2173 AGGCGGCGCAGCTCGGAGGGCCGGCGCTCGTCCTGCGGCAGCGACAGGGAATTCAGCGCC 2232

Qy        718 CGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGC 777

Db       2233 CGGTCCACCTCCGCCGTCGCGGCCTCCTCACCCAGCAGCGGGGCCAGCTTGCGGCGGAAC 2292

Qy        778 GCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTG 837

Db       2293 TCG---TCGGGCGAGCCCGCCACCACGCCGCCCGCGGCGGGCGCGGCCTCCTCGCGTGTG 2349

Qy        838 CACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCG 897

Db       2350 CAGGCCCGCGCGGCCTCTGCTTCCTCGGCGGACTGGCGCGTCGCCAGCGACACGCCGATG 2409

Qy        898 AGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGC 957

Db       2410 AAGGCCAGCCCGTTGAGCGTGAGCGAGGCGAAGGTGGCGAAGCCCCAGGGCTCACTGTCG 2469

Qy        958 GCCGCGCCCGCCGGCGCGCCCGCGC 982

Db       2470 GAGAACCCGAGCAGCGCCGCCATGC 2494
```

```
RESULT 3
US-08-911-853-28/c
; Sequence 28, Application US/08911853
; Patent No. 6048710
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/911,853
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/699,092
;      FILING DATE:  16-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-911-853-28

  Query Match             17.4%;  Score 174.2;  DB 3;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 9.7e-19;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||| || ||||| |||| |||||     |  | || ||||      |    ||||  ||
Db        632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |   ||   |||| ||   ||| |||   |||  |    ||  |    |
Db        575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                | |  || ||| | |   | || ||                  | ||||||     |
Db        515 TGCTCCTCGTCCTGGGCGATGAAGTCC---------------TGCAGCAGCTCGCGCCAG 471

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||  |  ||| | |   ||  |||  ||||| |  |   || |   |  ||||  |||
Db        470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||  | |  |||||    |     |   |   |||   | |||  |  || || ||
Db        410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || | |   ||   |  |||  |  | || |||||    |||||||    |||
Db        350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  ||| |||| || |      | | |||||  | ||| |   |  |||  || |||
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              |||    | |      |||| |     |  | |   |  |  || ||   |||   |||
Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              || | |  || | |||||          || || |||   |  |  || ||  ||||
Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | | ||  ||  ||| ||   |||||||   |  | ||  || ||| | ||||  |||
Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
              || ||  ||   || | | |   ||| ||| ||   || |||  | |  | |
Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 4
US-09-479-409-28/c
; Sequence 28, Application US/09479409
; Patent No. 6225106
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,409
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-409-28

  Query Match             17.4%;  Score 174.2;  DB 3;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 9.7e-19;
```

```
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              |||||| || ||||| |||| |||||     |  | || ||||      |    ||||  ||
Db        632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |   ||   |||| ||   ||| |||   |||  |    ||  |    |
Db        575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                | |  || ||| | |   | || ||                 | ||||||      |
Db        515 TGCTCCTCGTCCTGGGCGATGAAGTCC---------------TGCAGCAGCTCGCGCCAG 471

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||  |  ||| | |   || |||  ||||| |  |   || |   |  ||||  |||
Db        470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              |||| | |   |||||    |      |   |   |||   | |||  |  || || ||
Db        410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || | |   ||   | |||  |  | || |||||    ||||||||    |||
Db        350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  ||| |||| || |     | | |||||  |  |||  |   |    |||  || |||
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              |||     | |       |||| |       | | |   | |  || ||   |||   |||
Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              || | |  || | |||||          || || |||   |  |  || ||  ||||
Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              | | ||  ||  ||| ||   |||||||    |  | ||  || |||  | |||| |||
Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
                || ||  ||   || | | |   ||| ||| ||   || |||  | |  | |
Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 5
US-09-479-453-28/c
; Sequence 28, Application US/09479453
; Patent No. 6313283
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,453
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-453-28

  Query Match             17.4%;  Score 174.2;  DB 3;  Length 4377;
  Best Local Similarity   54.8%;  Pred. No. 9.7e-19;
  Matches  425;  Conservative    0;  Mismatches  323;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        752 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 693

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        692 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 633

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              |||||| || ||||| |||| |||||     |  | || ||||      |    ||||  ||
Db        632 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 576

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  |||||  |   ||   |||| ||   ||| |||   |||  |    ||  |    |
Db        575 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 516

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                | |  || ||| | |   | || ||                 | ||||||      |
Db        515 TGCTCCTCGTCCTGGGCGATGAAGTCC---------------TGCAGCAGCTCGCGCCAG 471

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |||  |  ||| | |   || |||  ||||| |  |   || |   |  ||||  |||
Db        470 GGGTGCTCCAGGCTGAGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCGGTGAGT 411

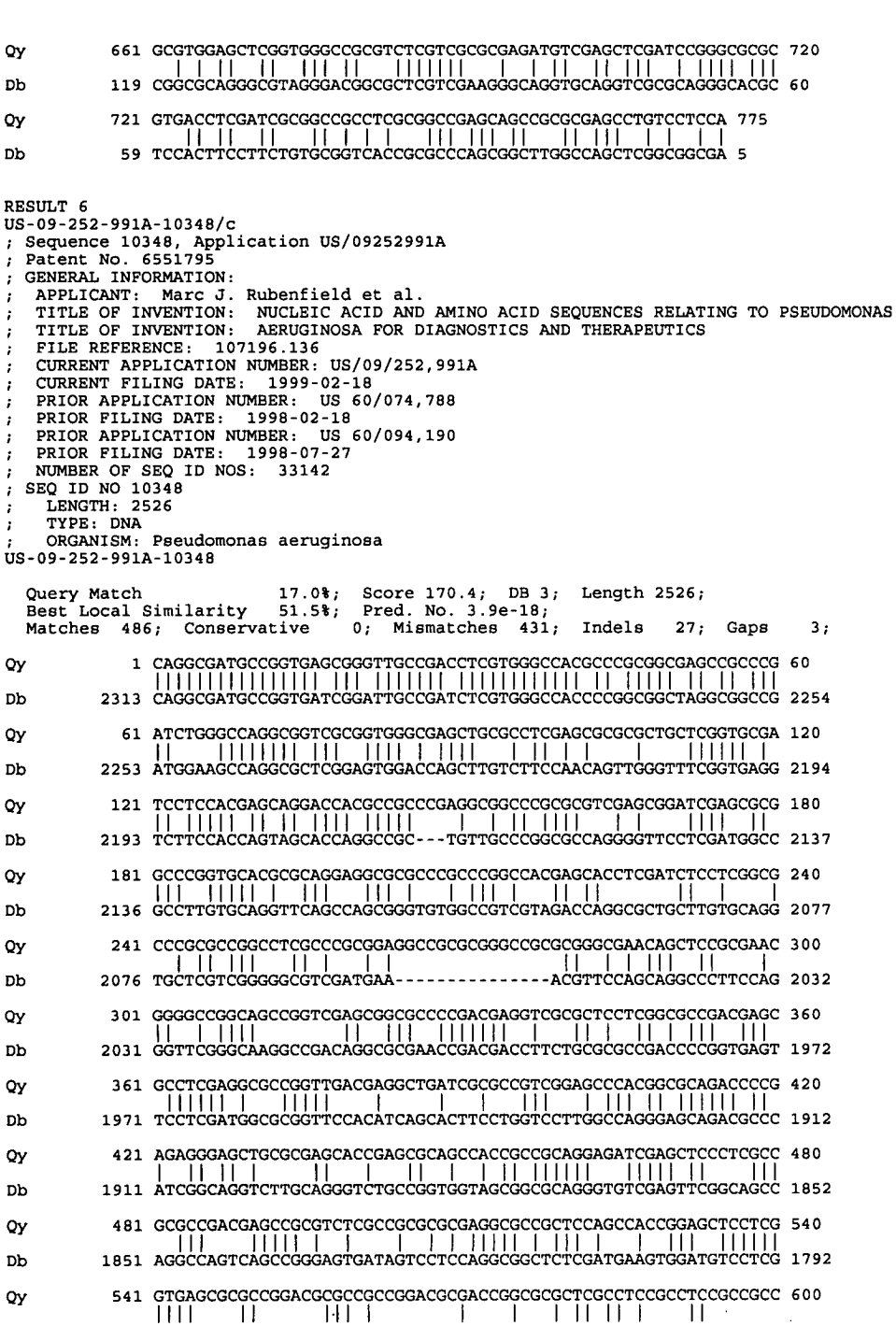
Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              |||| | |   |||||    |      |   |   |||   | |||  |  || || ||
Db        410 TCCTCCATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGCGAACATACGCCC 351

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  || || | |   ||   | |||  |  | || |||||    ||||||||    |||
Db        350 ATCGGCAGTTCCTGCAGGGTCTGGCGGTGGAAGCGGCGCAGCGCGTCGAGCTCGGCGGCC 291

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||  ||| |||| || |     | | |||||  |  |||  |   |    |||  || |||
Db        290 AGGCCGGTGAGGCGCGACTGGTAATCCTCCAGGCGACTCTCGATGAAGTGGATGTCTTCG 231

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              |||     | |       |||| |       | | |   | |  || ||   |||   |||
Db        230 CTGACATAGGCCTCGCTGCCGGCCTTGTAGGGCAGGAAGGTTTCCACCATGTCCTGGGCC 171

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              || | |  || | |||||          || || |||   |  |  || ||  ||||
Db        170 ACGCTGGGGCCCATCAGGCCG---------GAGAGATTGGCCTCGATGCGGTCGCGCAGA 120
```



```
Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
Db        119 CGGCGCAGGGCGTAGGGACGGCGCTCGTCGAAGGGCAGGTGCAGGTCGCGCAGGGCACGC 60

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCA 775
Db         59 TCCACTTCCTTCTGTGCGGTCACCGCGCCCAGCGGCTTGGCCAGCTCGGCGGCGA 5


RESULT 6
US-09-252-991A-10348/c
; Sequence 10348, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10348
;   LENGTH: 2526
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10348

  Query Match             17.0%;  Score 170.4;  DB 3;  Length 2526;
  Best Local Similarity   51.5%;  Pred. No. 3.9e-18;
  Matches  486;  Conservative    0;  Mismatches  431;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
Db       2313 CAGGCGATGCCGGTGATCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCTAGGCGGCCG 2254

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
Db       2253 ATGGAAGCCAGGCGCTCGGAGTGGACCAGCTTGTCTTCCAACAGTTGGGTTTCGGTGAGG 2194

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
Db       2193 TCTTCCACCAGTAGCACCAGGCCGC---TGTTGCCCGGCGCCAGGGGTTCCTCGATGGCC 2137

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
Db       2136 GCCTTGTGCAGGTTCAGCCAGCGGGTGTGGCCGTCGTAGACCAGGCGCTGCTTGTGCAGG 2077

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
Db       2076 TGCTCGTCGGGGGCGTCGATGAA--------------ACGTTCCAGCAGGCCCTTCCAG 2032

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
Db       2031 GGTTCGGGCAAGGCCGACAGGCGCGAACCGACGACCTTCTGCGCGCCGACCCCGGTGAGT 1972

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
Db       1971 TCCTCGATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGGGAGCAGACGCCC 1912

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
Db       1911 ATCGGCAGGTCTTGCAGGGTCTGCCGGTGGTAGCGGCGCAGGGTGTCGAGTTCGGCAGCC 1852

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
Db       1851 AGGCCAGTCAGCCGGGAGTGATAGTCCTCCAGGCGGCTCTCGATGAAGTGGATGTCCTCG 1792

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
Db       1791 GTGACATAGCTTTCTTCGCTGGTCTTGTAGGGAAGGAAGGTCTCCACGATGTC------- 1739

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
Db       1738 --CTGGGCGACACTGGGTCCCATCAGCCCGGAAAGGTTCGCCTCGATGCGGTCGCGCAAG 1681

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
Db       1680 CGACGCAGGGCATAGGGGCGACGTTCGTCGAAGGGCAGGTGGAGATCGCGCAGGGCCTGC 1621

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
Db       1620 TCGACCTCCTTCTGCGCGGTCTTGGCGCCCAGCGGCTTGGCCAACTGGCTGGCGAACTCC 1561

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
Db       1560 TGCGGCGAACCGGCGAACAGTTCGCGGCGTTGCGGACGGCGTACGTTGTCCACCGCGCAG 1501

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
Db       1500 GCCTCGGCGGCGCCTTTTTCCTCGTCGCTGGCCTCGGTGAACAGCGACACCAGGGTGAAT 1441

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
Db       1440 ACCAGCACGTTGGCCGCCAGCGAGGACAGCGCCGCCAGGTGCCA 1397


RESULT 7
US-09-252-991A-10447/c
; Sequence 10447, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10447
;   LENGTH: 2991
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10447

  Query Match             17.0%;  Score 170.4;  DB 3;  Length 2991;
  Best Local Similarity   51.5%;  Pred. No. 3.8e-18;
  Matches  486;  Conservative    0;  Mismatches  431;  Indels   27;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
Db       2369 CAGGCGATGCCGGTGATCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCTAGGCGGCCG 2310

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
Db       2309 ATGGAAGCCAGGCGCTCGGAGTGGACCAGCTTGTCTTCCAACAGTTGGGTTTCGGTGAGG 2250

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
Db       2249 TCTTCCACCAGTAGCACCAGGCCGC---TGTTGCCCGGCGCCAGGGGTTCCTCGATGGCC 2193

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
Db       2192 GCCTTGTGCAGGTTCAGCCAGCGGGTGTGGCCGTCGTAGACCAGGCGCTGCTTGTGCAGG 2133

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
Db       2132 TGCTCGTCGGGGGCGTCGATGAA--------------ACGTTCCAGCAGGCCCTTCCAG 2088
```

```
Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
Db       2087 GGTTCGGGCAAGGCCGACAGGCGCGAACCGACGACCTTCTGCGCGCCGACCCCGGTGAGT 2028

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
Db       2027 TCCTCGATGGCGCGGTTCCACATCAGCACTTCCTGGTCCTTGGCCAGGGAGCAGACGCCC 1968

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
Db       1967 ATCGGCAGGTCTTGCAGGGTCTGCCGGTGGTAGCGGCGCAGGGTGTCGAGTTCGGCAGCC 1908

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
Db       1907 AGGCCAGTCAGCCGGGAGTGATAGTCCTCCAGGCGGCTCTCGATGAAGTGGATGTCCTCG 1848

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
Db       1847 GTGACATAGCTTTCTTCGCTGGTCTTGTAGGGAAGGAAGGTCTCCACGATGTC------- 1795

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
Db       1794 --CTGGGCGACACTGGGTCCCATCAGCCCGGAAAGGTTCGCCTCGATGCGGTCGCGCAAG 1737

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
Db       1736 CGACGCAGGGCATAGGGGCGACGTTCGTCGAAGGGCAGGTGGAGATCGCGCAGGGCCTGC 1677

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
Db       1676 TCGACCTCCTTCTGCGCGGTCTTGGCGCCCAGCGGCTTGGCCAACTGGCTGGCGAACTCC 1617

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
Db       1616 TGCGGCGAACCGGCGAACAGTTCGCGGCGTTGCGGACGGCGTACGTTGTCCACCGCGCAG 1557

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
Db       1556 GCCTCGGCGGCGCCTTTTTCCTCGTCGCTGGCCTCGGTGAACAGCGACACCAGGGTGAAT 1497

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCA 944
Db       1496 ACCAGCACGTTGGCCGCCAGCGAGGACAGCGCCGCCAGGTGCCA 1453
```

```
RESULT 8
US-08-911-853-1/c
; Sequence 1, Application US/08911853
; Patent No. 6048710
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/911,853
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/699,092
;      FILING DATE:  16-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1029 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-911-853-1

  Query Match             11.0%;  Score 110.2;  DB 3;  Length 1029;
  Best Local Similarity   59.9%;  Pred. No. 6.9e-09;
  Matches  203;  Conservative    0;  Mismatches  133;  Indels    3;  Gaps    1;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
Db        404 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 345

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
Db        344 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 285

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
Db        284 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 228

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
Db        227 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 168

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
Db        167 TGCTCCTCGTCCTGGGCGATGAAGTCCTGCAGCAGCTCGCGCCAGGGGTGCTCCAGGCTG 108

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCG 339
Db        107 AGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCG 69
```

```
RESULT 9
US-09-479-409-1/c
; Sequence 1, Application US/09479409
; Patent No. 6225106
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,409
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1029 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-409-1

  Query Match             11.0%;  Score 110.2;  DB 3;  Length 1029;
  Best Local Similarity   59.9%;  Pred. No. 6.9e-09;
  Matches  203;  Conservative    0;  Mismatches  133;  Indels    3;  Gaps    1;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              |||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        404 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 345

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        344 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 285

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||| || ||||| |||| |||||    |  | || ||||     |    ||||  ||
Db        284 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 228

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |   ||   |||| |||   |||  |   |||  |    |  |    |
Db        227 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 168

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                | |  || ||| | |   | || ||    |   | ||||   |    |||||  |
Db        167 TGCTCCTCGTCCTGGGCGATGAAGTCCTGCAGCAGCTCGCGCCAGGGGTGCTCCAGGCTG 108

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCG 339
                |    ||   | | | | | ||  | ||  || | ||
Db        107 AGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCG 69


RESULT 10
US-09-479-453-1/c
; Sequence 1, Application US/09479453
; Patent No. 6313283
;  GENERAL INFORMATION:
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Quax, Wilhelmus J.
;    TITLE OF INVENTION:  EXPRESSION SYSTEM FOR ALTERED
;    TITLE OF INVENTION:  EXPRESSION LEVELS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,453
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/911,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Glaister, Debra J
;      REGISTRATION NUMBER:  33,888
;      REFERENCE/DOCKET NUMBER:  GC361-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-846-7620
;      TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1029 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-479-453-1

  Query Match             11.0%;  Score 110.2;  DB 3;  Length 1029;
  Best Local Similarity   59.9%;  Pred. No. 6.9e-09;
  Matches  203;  Conservative    0;  Mismatches  133;  Indels    3;  Gaps    1;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              |||||||||||||||| ||| ||||||| |||||||||||| || ||||| || || |||
Db        404 CAGGCGATGCCGGTGACCGGATTGCCGATCTCGTGGGCCACCCCGGCGGCCAGGCGGCCG 345

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||   |||||| || |||  |||  | |||||  | || ||| ||||   |||||||
Db        344 ATGCTGGCCAGACGCTCGGAGTGCACCAGCTGGTCTTCCAGCACGCGGGTCTCGGTGACG 285

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||| || ||||| |||| |||||    |  | || ||||     |    ||||  ||
Db        284 TCCTCGACCAGCAGCACCAGGCCGC---TGTTGCCCGGCGCCAGCGGTTCGTCGATGGCC 228

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              |||  ||||| |   ||   |||| |||   |||  |   |||  |    |  |    |
Db        227 GCCTTGTGCAGGTTGAGCCAGCGCACCTCGCCGTCCAGTTGCAGGTGCTGCTTGTGCAGG 168

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
                | |  || ||| | |   | || ||    |   | ||||   |    |||||  |
Db        167 TGCTCCTCGTCCTGGGCGATGAAGTCCTGCAGCAGCTCGCGCCAGGGGTGCTCCAGGCTG 108

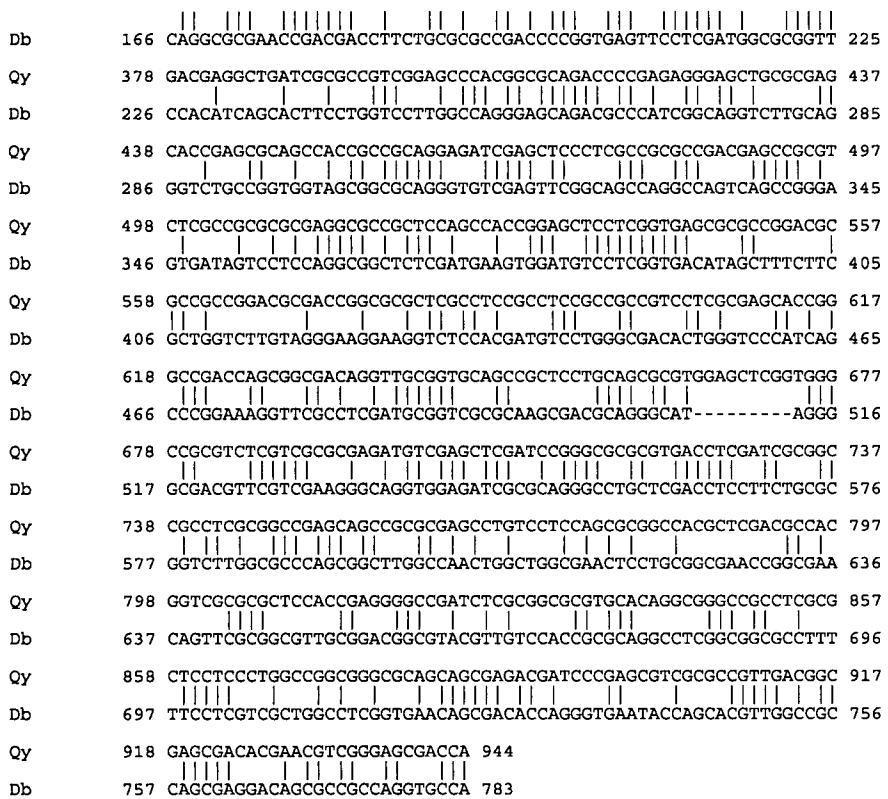
Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCG 339
                |    ||   | | | | | ||  | ||  || | ||
Db        107 AGCAGGCGCGAGCCGACCACCTGCTGCGCGCTGATGCCG 69


RESULT 11
US-09-252-991A-10720
; Sequence 10720, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10720
;   LENGTH: 2217
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10720

  Query Match             10.5%;  Score 105.4;  DB 3;  Length 2217;
  Best Local Similarity   48.3%;  Pred. No. 3.6e-08;
  Matches  332;  Conservative    0;  Mismatches  346;  Indels    9;  Gaps    1;

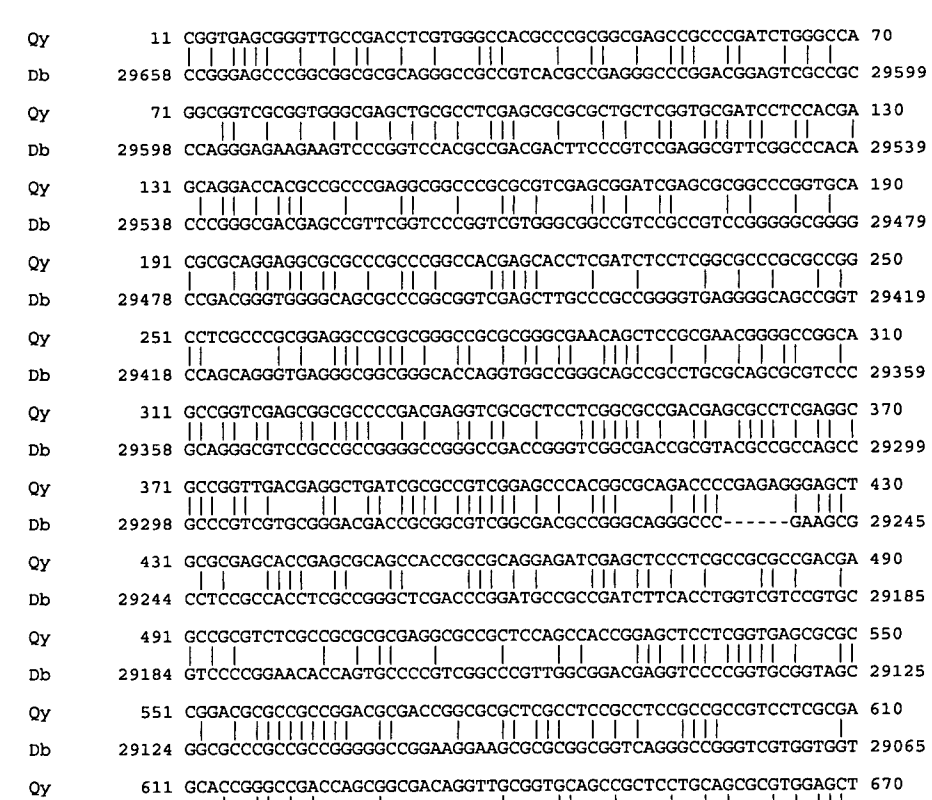
Qy        258 CGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTC 317
              | ||  ||  |||  |       ||  | | |||  ||    | ||  | ||||
Db        106 CTCGTCGGGGGCGTCGATGAAACGTTCCAGCAGGCCCTTCCAGGGTTCGGGCAAGGCCGA 165

Qy        318 GAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTT 377
```



RESULT 12
US-09-477-962-1/c
; Sequence 1, Application US/09477962
; Patent No. 6927286
; GENERAL INFORMATION:
; APPLICANT: SHEN, BEN
; APPLICANT: DU, LIANGCHENG
; APPLICANT: SANCHEZ, CESAR
; APPLICANT: CHEN, MEI
; APPLICANT: EDWARDS, DANIEL J.
; TITLE OF INVENTION: BLEOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES
; FILE REFERENCE: 407T-895820US
; CURRENT APPLICATION NUMBER: US/09/477,962
; CURRENT FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: 60/115,435
; PRIOR FILING DATE: 1999-01-06
; PRIOR APPLICATION NUMBER: 60/118,848
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1
; LENGTH: 58857
; TYPE: DNA
; ORGANISM: Streptomyces verticillus
US-09-477-962-1

Query Match 10.5%; Score 105.4; DB 3; Length 58857;
Best Local Similarity 45.0%; Pred. No. 2.8e-08;
Matches 443; Conservative 0; Mismatches 536; Indels 6; Gaps 1;



RESULT 13
US-09-902-540-8819/c
; Sequence 8819, Application US/09902540

```
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8819
;   LENGTH: 6645
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-8819

  Query Match             10.2%;  Score 102.4;  DB 3;  Length 6645;
  Best Local Similarity   45.9%;  Pred. No. 9.4e-08;
  Matches  432;  Conservative    0;  Mismatches  501;  Indels    9;  Gaps    2;

Qy         48 GGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCG 107
Db       1083 GGCGCGCTCGTCCTTCTCGACCAGCGTGGCGGACAGGGCCAGCAGCCGCTCGCGAGCGCG 1024

Qy        108 CTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAG 167
Db       1023 GGCGGCGAGGTTGGCGTCGGCGGATACGGCGCGGCGCCACACGTCCAGCTCCGCGGCCTC 964

Qy        168 CGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTC 227
Db        963 ATCGCCCGACTCGCGCGCCAGCTCCGTCAGCGCGAGCAGCAACGAGGCGGGGCGACGAGA 904

Qy        228 GATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAA 287
Db        903 ACCCTTCGCGGCCTCCGAGAAGTCATCGCGCGCGGCGGCCACCCGGCCCGCGGCGCGGTG 844

Qy        288 CAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTC 347
Db        843 AAGCACGGCGCGCCGGGCCAGCAGC---TCCGCGCGCGCCTCGCCCGAGGCGGCGGCCAC 787

Qy        348 GGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCAC 407
Db        786 GAGCGTGCTCAGCAACTCCAGCCGCTCCGTCTCCTCTTCCGGCGTGCCGGTGCCGGGTGG 727

Qy        408 GGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATC 467
Db        726 AGGCAGCAGGTCCAGCAGCGACTGCGCCGCCCAGGTGTCTTGCGGCTCCTCCAG------ 673

Qy        468 GAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCA 527
Db        672 GCGCAGGCTGCGCAGCGCGTCCACCGCGGAGGTGGCGCGCCCCAGGCCCAGCGACAGCTC 613

Qy        528 CCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTC 587
Db        612 GGCCAGCTCGCGGCGGGCGGCGCGGCGCGGCTCGCCGGACAGGCGCGGCCACAGCTCCAC 553

Qy        588 CGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAG 647
Db        552 CAGCAGGTCCGCCAGGGCCTCGCGCGCACCCGCCTTGCGGTGCAGCGACGCCAGCTCCAC 493

Qy        648 CCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTC 707
Db        492 CCGGGCGTCCAGATCTTCCGGCGCGCGGGCGCTGTACTCGGCCAGCAGGCGGCGGGCGGC 433

Qy        708 GATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCT 767
Db        432 ATCCGTGCGGCTCGCGCGCACCGCCGCGGTGGCGGCCTTGCGCGTCAGGGCTACGCGCTC 373

Qy        768 GTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTC 827
Db        372 ACTCTCGCGAGGCGCATCGCCGTCCGCGGGCGGCAGCGCCAGCACGGCCTCGAAGTCCTC 313

Qy        828 GCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGA 887
Db        312 CAAAGCGGCGCGGGCATCCGTGGGCAGACGCAGCTCGCCCCGGCGCTGGCGCGCGGGCGC 253

Qy        888 GACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGG 947
Db        252 GTAGCTCGCGTCACGCTCCAGGGCGTCCGCGAGGAAGGCCTCCTCGCGGCTGGTCCCCGC 193

Qy        948 GTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCCGCG 989
Db        192 GGCGAGCAGCGAGGCGCGGAACAGCAGCTTCGCGCCAGCGCG 151


RESULT 14
US-09-902-540-938/c
; Sequence 938, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 938
;   LENGTH: 10210
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-938

  Query Match             10.2%;  Score 102.4;  DB 3;  Length 10210;
  Best Local Similarity   45.9%;  Pred. No. 9.1e-08;
  Matches  432;  Conservative    0;  Mismatches  501;  Indels    9;  Gaps    2;

Qy         48 GGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCG 107
Db       1458 GGCGCGCTCGTCCTTCTCGACCAGCGTGGCGGACAGGGCCAGCAGCCGCTCGCGAGCGCG 1399

Qy        108 CTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAG 167
Db       1398 GGCGGCGAGGTTGGCGTCGGCGGATACGGCGCGGCGCCACACGTCCAGCTCCGCGGCCTC 1339

Qy        168 CGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTC 227
Db       1338 ATCGCCCGACTCGCGCGCCAGCTCCGTCAGCGCGAGCAGCAACGAGGCGGGGCGACGAGA 1279

Qy        228 GATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAA 287
Db       1278 ACCCTTCGCGGCCTCCGAGAAGTCATCGCGCGCGGCGGCCACCCGGCCCGCGGCGCGGTG 1219

Qy        288 CAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTC 347
Db       1218 AAGCACGGCGCGCCGGGCCAGCAGC---TCCGCGCGCGCCTCGCCCGAGGCGGCGGCCAC 1162

Qy        348 GGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCAC 407
Db       1161 GAGCGTGCTCAGCAACTCCAGCCGCTCCGTCTCCTCTTCCGGCGTGCCGGTGCCGGGTGG 1102

Qy        408 GGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATC 467
Db       1101 AGGCAGCAGGTCCAGCAGCGACTGCGCCGCCCAGGTGTCTTGCGGCTCCTCCAG------ 1048

Qy        468 GAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCA 527
Db       1047 GCGCAGGCTGCGCAGCGCGTCCACCGCGGAGGTGGCGCGCCCCAGGCCCAGCGACAGCTC 988
```

```
Qy       528 CCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTC 587
Db       987 GGCCAGCTCGCGGCGGGCGGCGCGGCGCGGCTCGCCGGACAGGCGCGGCCACAGCTCCAC 928

Qy       588 CGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAG 647
Db       927 CAGCAGGTCCGCCAGGGCCTCGCGCGCACCCGCCTTGCGGTGCAGCGACGCCAGCTCCAC 868

Qy       648 CCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTC 707
Db       867 CCGGGCGTCCAGATCTTCCGGCGCGCGGGCGCTGTACTCGGCCAGCAGGCGGCGGGCGGC 808

Qy       708 GATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCT 767
Db       807 ATCCGTGCGGCTCGCGCGCACCGCCGCGGTGGCGGCCTTGCGCGTCAGGGCTACGCGCTC 748

Qy       768 GTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTC 827
Db       747 ACTCTCGCGAGGCGCATCGCCGTCCGCGGGCGGCAGCGCCAGCACGGCCTCGAAGTCCTC 688

Qy       828 GCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGA 887
Db       687 CAAAGCGGCGCGGGCATCCGTGGGCAGACGCAGCTCGCCCCGGCGCTGGCGCGCGGGCGC 628

Qy       888 GACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGG 947
Db       627 GTAGCTCGCGTCACGCTCCAGGGCGTCCGCGAGGAAGGCCTCCTCGCGGCTGGTCCCCGC 568

Qy       948 GTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCCGCG 989
Db       567 GGCGAGCAGCGAGGCGCGGAACAGCAGCTTCGCGCCAGCGCG 526


RESULT 15
US-09-758-759-150/c
; Sequence 150, Application US/09758759
; Patent No. 6861513
; GENERAL INFORMATION:
;  APPLICANT: Hosted, Thomas J.
;  APPLICANT:  Wang, Tim X.
;  APPLICANT:  Horan, Ann C.
;  TITLE OF INVENTION: Everninomicin Biosynthetic Genes
;  FILE REFERENCE: ID0983K US
;  CURRENT APPLICATION NUMBER: US/09/758,759
;  CURRENT FILING DATE:  2001-01-11
;  PRIOR APPLICATION NUMBER: US 60/175,751
;  PRIOR FILING DATE: 2000-01-12
;  NUMBER OF SEQ ID NOS: 204
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 150
;   LENGTH: 1839
;   TYPE: DNA
;   ORGANISM: Micromonospora carbonacea
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1839)
;   OTHER INFORMATION: evcE
US-09-758-759-150

  Query Match            10.2%;  Score 101.8;  DB 3;  Length 1839;
  Best Local Similarity   45.4%;  Pred. No. 1.3e-07;
  Matches  450;  Conservative    0;  Mismatches  532;  Indels    9;  Gaps    2;

Qy         3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
Db      1218 GGCGAGGGTGCTCTTGCCGGCGCCGCTCGGCCCGACCACGGCGACCCGGCGGCCGGCCGG 1159

Qy        63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATC 122
Db      1158 CAGGTCCAGGGTGACCCGGTCCAGGGCCGGGGCGGTGCCGGCCCGGTACCGCACGGTGAC 1099

Qy       123 CTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGC 182
Db      1098 CGCGTCGAAGCGCACGTCGTGCGGGCCGGCACCGACGGCGGCGGCACGGGCGGCACCGGG 1039

Qy       183 CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCC 242
Db      1038 CGGGGTGGCGGCGGGCGCGTCGGCCTGCGGGGCGGTCAGCAGGGCGGCCACCCGGACCAG 979

Qy       243 CGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGG 302
Db       978 CCCGGCCCGGAGCTGGGTGCGCTGCCGGGCGGCCCCGACCAGCGCCAGCGCCACCTCGAC 919

Qy       303 GGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGC 362
Db       918 GGCGGCCAGGGAACCGACCGCCAGCACCCCGACCAGCACCCCGCCGACGCCGTCGCGCAG 859

Qy       363 CTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAG 422
Db       858 CGCGGTGACCACCACCGTGCCGGCGGTCACCCCGGCGACGAGCGCCCCGGCGGCGTCCAC 799

Qy       423 AGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGC 482
Db       798 GGCGAAGCCGGTGGCGGCGAGCCGTCGTTCCAGCCGGGCCAGCCGGCGGGCCCGATCGGC 739

Qy       483 GCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGT 542
Db       738 GGCGGCGTCCAGCGCGTACCCGGTGGCACCGAACGCGGCCAGGTCGGCGGCGCCGTGGAC 679

Qy       543 GAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGT 602
Db       678 AAGGTCCACGGCGTCCGTGGCGAGCGCGCCGCGCAGCGGGGCCACCCGGTCGGCGGCGTG 619

Qy       603 CCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGC 662
Db       618 CCGGGTCAGCGCGGTGGCCGCGAGCGGCAGGGCCACCCCGGCGACGAGCAGGCCCAGCGC 559

Qy       663 GTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGT 722
Db       558 CAGCACCCCGGCGGCGGGGAGCGAGATGGTGGTGGCCCCGGCCACGGCCAGCACGC---T 502

Qy       723 GACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGC 782
Db       501 GACCACCGTGGCGGCGGCCCCCGGGACGAGCACCCGCAGCAGCAGGTCCTGCACGGCCTC 442

Qy       783 CACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAG 842
Db       441 CACGTCGGACACGAGCCGGCTCAGCGCGTCCCCGGTGCGCTGCCGGGCGGCGTCGCGCCG 382

Qy       843 GCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGT 902
Db       381 GGCGGCCAGGGCGGCGAAGACCCCGGCC------CGGACGTCGGTGATCATCCGCAGCAC 328

Qy       903 CGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGC 962
Db       327 GGCATCGTGGCCGGCGAGGCGCTCGGTGTAGCGGAACACGCCTCGGCTGATCGCCAGCGC 268

Qy       963 GCCCGCCGGCGCGCCCGCGCCCCCGCGCAGG 993
Db       267 CCGGACCGCGACGATCGCCACGGTGAGCCGG 237


Search completed: April 27, 2006, 12:42:32
Job time : 208.521 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 1003.57 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_2000_3000
Perfect score: 1001
Sequence: 1 caggcgatgccggtgagcgg.....cccccgcgcaggagcgcgag 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1001	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
c	2	111	11.1	3513	7	US-10-282-122A-13820	Sequence 13820, A
c	3	109.6	10.9	3459	6	US-10-369-493-31552	Sequence 31552, A
	4	107.6	10.7	1172	8	US-10-363-345A-18585	Sequence 18585, A
c	5	107.6	10.7	1172	8	US-10-363-345A-18586	Sequence 18586, A
c	6	107.6	10.7	1172	8	US-10-363-345A-18587	Sequence 18587, A
	7	107.6	10.7	1172	8	US-10-363-345A-18588	Sequence 18588, A
	8	107.6	10.7	1172	9	US-10-363-483A-18585	Sequence 18585, A
c	9	107.6	10.7	1172	9	US-10-363-483A-18586	Sequence 18586, A
c	10	107.6	10.7	1172	9	US-10-363-483A-18587	Sequence 18587, A
	11	107.6	10.7	1172	9	US-10-363-483A-18588	Sequence 18588, A
c	12	107	10.7	1172	8	US-10-363-345A-18585	Sequence 18585, A
	13	107	10.7	1172	8	US-10-363-345A-18586	Sequence 18586, A
	14	107	10.7	1172	8	US-10-363-345A-18587	Sequence 18587, A
c	15	107	10.7	1172	8	US-10-363-345A-18588	Sequence 18588, A
c	16	107	10.7	1172	9	US-10-363-483A-18585	Sequence 18585, A
	17	107	10.7	1172	9	US-10-363-483A-18586	Sequence 18586, A
	18	107	10.7	1172	9	US-10-363-483A-18587	Sequence 18587, A
c	19	107	10.7	1172	9	US-10-363-483A-18588	Sequence 18588, A
c	20	106.8	10.7	77294	8	US-10-729-802-1	Sequence 1, Appli
c	21	106	10.6	1515	7	US-10-437-963-97778	Sequence 97778, A
c	22	105.8	10.6	6408	6	US-10-329-079-44	Sequence 44, Appl
c	23	105.8	10.6	61944	6	US-10-329-079-34	Sequence 34, Appl
	24	105.6	10.5	138203	9	US-10-819-386A-1	Sequence 1, Appli
c	25	104.6	10.4	1236	7	US-10-437-963-98599	Sequence 98599, A
c	26	104.4	10.4	88421	3	US-09-976-059-1	Sequence 1, Appli
c	27	104	10.4	11040	10	US-11-109-593-3	Sequence 3, Appli
c	28	104	10.4	67323	10	US-11-109-593-1	Sequence 1, Appli
c	29	103.8	10.4	10002	7	US-10-282-122A-14471	Sequence 14471, A
	30	103.2	10.3	1476	7	US-10-437-963-42144	Sequence 42144, A
c	31	102.8	10.3	18876	6	US-10-329-079-42	Sequence 42, Appl
c	32	102.6	10.2	1677	6	US-10-156-761-3331	Sequence 3331, Ap
	33	102.6	10.2	9025608	6	US-10-156-761-1	Sequence 1, Appli
c	34	102.4	10.2	15738	6	US-10-329-079-46	Sequence 46, Appl
c	35	102	10.2	3024	9	US-10-939-036-35	Sequence 35, Appl
c	36	101.8	10.2	1839	3	US-09-758-759-150	Sequence 150, App
c	37	101.8	10.2	1839	10	US-11-021-825-150	Sequence 150, App
c	38	101.8	10.2	5643	10	US-11-109-593-5	Sequence 5, Appli
c	39	101.8	10.2	109519	3	US-09-758-759-1	Sequence 1, Appli
c	40	101.8	10.2	109519	10	US-11-021-825-1	Sequence 1, Appli
	41	101.6	10.1	3024	9	US-10-939-036-35	Sequence 35, Appl
c	42	101.2	10.1	77536	3	US-09-940-316B-1	Sequence 1, Appli
c	43	100.6	10.0	6432	6	US-10-329-079-10	Sequence 10, Appl
c	44	100.6	10.0	37360	6	US-10-329-079-6	Sequence 6, Appli
c	45	100.2	10.0	4725	5	US-10-205-032-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 1001; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 1.5e-197;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2000 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 2059

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2060 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 2119

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2120 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 2179
```

```
Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2180 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 2239

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2240 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 2299

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2300 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 2359

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2360 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 2419

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2420 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 2479

Qy        481 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2480 GCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCG 2539

Qy        541 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2540 GTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCC 2599

Qy        601 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2600 GTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGC 2659

Qy        661 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2660 GCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGC 2719

Qy        721 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2720 GTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCG 2779

Qy        781 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2780 GCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCAC 2839

Qy        841 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2840 AGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGC 2899

Qy        901 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2900 GTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCC 2959

Qy        961 GCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001
              |||||||||||||||||||||||||||||||||||||||||
Db       2960 GCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 3000


RESULT 2
US-10-282-122A-13820/c
; Sequence 13820, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13820
;   LENGTH: 3513
;   TYPE: DNA
;   ORGANISM: Burkholderia mallei
US-10-282-122A-13820

  Query Match            11.1%;  Score 111;  DB 7;  Length 3513;
  Best Local Similarity   48.1%;  Pred. No. 2.7e-14;
  Matches  476;  Conservative    0;  Mismatches  495;  Indels   18;  Gaps    5;

Qy         26 CGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGG 85
              || |||||| | || || |||  |||||| || ||       |  ||         | |
Db       2815 CGGCCTCGTCGACCTCGGCCGTCGCGAGCTGCTCGGCGAACTGCTCGCCCGTCATACGCG 2756

Qy         86 CGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGC 145
              |   |||| |||  ||     |||    | |||||||   ||   || | | |  ||
Db       2755 CCGCCTGCTCCTTCAGTTGCAGCTCGGTGATGCGATCGCGCAGCGGCTGCAGCGAGCGCT 2696

Qy        146 CCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCG 205
              | | ||||  |||||| |||   |   || ||      |       |       |||   ||
Db       2695 CGGCGGCGAGCCGCGCCTCGTCGGCCGCGCGCAGCTTCGCCGTGAGATCGTCGAGCTCCG 2636

Qy        206 CCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGG 265
              | ||| ||||  ||||| |||    || | | |||| | |  ||  | || ||      |
Db       2635 CGCGCGCGGCGCCGAGCGCCTGCTCCTTCGCCGCGCGCACTTCGAGCGCGTCCTGCAGGC 2576

Qy        266 CCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCG 325
              | | | | || |   |  ||   | |  | ||| |   ||  ||      |||||| ||
Db       2575 CGGTGTGCGCGGTCTGCTCGTTGATCGTCTCGAGCTCCGCGCGCGCATCCTCGAGCGACG 2516

Qy        326 CCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGC 385
              || ||||| | |||  || ||||    ||||  |   |  | |   | | | | ||| ||
Db       2515 CCGCGACGCGCTCGGCCTGCTCGTGCGCGACCTGGATCGAGCGCTTCAGCTCGTCGATGC 2456

Qy        386 TGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGC 445
              | |||||   |   || | ||||| ||  | ||   | | |           |  ||
Db       2455 GGTTCGCGGACT---GGCGCGCGGCGAAGCGCGCGTCGGTGGCCGCGCGCGCTCGCGCTCGC 2399

Qy        446 GCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGC 505
              ||  |  | | |||  |     |||  |    |          ||| |     ||    |
Db       2398 GCGCCTCCTGGCGCGCGTTCGTCAGCGTCTCGTCGAGCGATTCGAACGCGAGCTGATTGT 2339

Qy        506 GCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTG-AGCGCGCCGGACGCGCCGCCG 564
                | ||| ||||  || ||       ||||||  ||  |  |||| |   |  |||| |||
```

```
Db       2338 CCTCGAAGCGCGCCTGCAATTCGGCGAGCTCGGCGTCGTGGCGCTCGAAATTCGCCTCCG 2279

Qy        565 GACGCGACCGGCGCGCTCG-----CCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGC 619
              ||  ||  |||| ||     ||||   || |||  |  ||  | ||||  ||  ||
Db       2278 ACTCCGCGCGCAGCGCACGCTGCTCCTCGATCTGCGCGCCGATCTCCTCGAGTTCCTCGC 2219

Qy        620 CGACCAGCG------GCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGG 673
              || | |||       ||  |   | |||  |   ||  || |   |||  | | |  ||
Db       2218 GGATCTGCGTGCTGCGCTGCGTATAGCGCTCGTGCGCCTGCGTGAGCTTGAGCACGTCCA 2159

Qy        674 TGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCG 733
              |  || |||  |   |||||    |  |  |||||  | | ||||||     |    | ||
Db       2158 TCTGCAGCGCGTGCACGCGCTGCGTCGCACGCTCGGCCTGCGCGCGCACTTCGGTCAGCG 2099

Qy        734 CGGCCGCCTCGCGGCCGAGC---AGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCG 790
              |   || | |  |   | ||       ||||  |  | || ||  |   ||| ||   |
Db       2098 CCTGCGACGCCTGCGTGTGCGCGGCTTCCGCGCGGATCGCCGCCGCCTTCGCCTCGTCGG 2039

Qy        791 ACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCG 850
              ||    | | |        |||| | |    | |      |     | ||| ||
Db       2038 ACAGGAGCGCCTGCGCGCGCACCTGGCGCGTCAGGTTCTCGATTTCCTGCGCGCGCGCGA 1979

Qy        851 CCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGT 910
              |  || | ||| |  ||      |       |||   |||  |  | ||||  || |||
Db       1978 GCATGCCCGCCTGCTCGGAGTCGGCCGCATAGAGCTGCACGCCCGAGCGCGTGACGACGT 1919

Qy        911 TGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCG 970
              | | ||   |  |||||||    |  |  |  |   |  ||| |||| |||  ||||
Db       1918 GGCCCGCCTTCACGACGAACGCGCCGCCCTGCGGCAATTGCATGCGCGCGGCGAGCGCCT 1859

Qy        971 GCGCGCCCGCGCCCCCGCGCAGGAGCGCG 999
              |||||    || |  ||   |  | || |
Db       1858 GCGCGAGGTCGTCGGCGACGAACACCGTG 1830
```

RESULT 3
US-10-369-493-31552/c
; Sequence 31552, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31552
; LENGTH: 3459
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31552

Query Match 10.9%; Score 109.6; DB 6; Length 3459;
Best Local Similarity 46.1%; Pred. No. 5.2e-14;
Matches 403; Conservative 0; Mismatches 469; Indels 2; Gaps 1;

```
Qy          26 CGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGG 85
               ||  |||| | ||| || ||| || || |||| ||  | | ||    |  |   | |
Db        2803 CGGTCTCGAGCGCCTCGACCGTGGGGATCCGCTCGGGATCGACCGCCAGCGCCTCGAGCA 2744

Qy          86 CGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGC 145
                ||||||     || |   || |||| |     |||  |   |||   ||   |
Db        2743 ACTGCTGCGGTGTCCGCTCCGTCTCCTCGCGGATCCGCTCGGCCGCCAGCTGCAGCGCCT 2684

Qy         146 CCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCG 205
               | |  |||||    |||  |  |    |  | |||  |   |    || |  |     |
```

```
Db        2683 CGGTCGCGGCATCGGCGCGGGCCTCGGCCCGGGCGCGCGCCTCGCGGCTCTCGCCCGCCT 2624

Qy         206 CCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGG 265
               |||  ||||| || || |  |   || | ||||||      | ||||||  ||| ||
Db        2623 GACGCTCGGCCTCGCGCTCGGCCTCCTGCGCGGCGCGGAGCGCCGCCTCGGCCGAGGCCA 2564

Qy         266 CCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCG 325
               |||   ||  |||||||||     ||||| ||   |   |   |      | |||  |
Db        2563 GCGCATCGGAGGCGCGGGCGCGGCGCTCCTCGGCGGCCTCGATCGCTTCTGCCAGCTCCT 2504

Qy         326 CCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGC 385
               |||  |  | | ||  ||||||||     |  ||    |||||  |  ||    |  ||
Db        2503 CCCGGCGGGCGGCGATCTCCTCGGGCGCCTCGGCGGCCTCGCGCAGCGCCTCCTCGGTCT 2444

Qy         386 TGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGC 445
               |  |  ||  ||||      ||||||  |  |     |  |  || |   |  |  |
Db        2443 CGGCCTTGCGCTCGGCCAGTTCGGCGCTGCGCTTCTCGGCGGTCTCCAGCCGGTGCTTCC 2384

Qy         446 GCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGC 505
               | | |  |  |  |  |  || || |  |  | |  |     |||| ||
Db        2383 AGCCCGAAAGGTCCTTCGTCACCTCCTGCCGCCGCTTGACCCGCGCCTCGCCCTCGCGGC 2324

Qy         506 GCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGG 565
               ||     |   ||  |  |  ||   |  |  | ||      ||| | || |  |
Db        2323 GCAGCTCGTCATGCAGCGAGCGGCGCGACATCATCGCGATGCGCGCGGCCTCGACCGCCA 2264

Qy         566 ACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCA 625
               |  |   | ||  | || |   | |||||  |      ||  ||| ||| | | ||
Db        2263 TCTTCAGATCCTCGAGCCCCGCGCGCGCCGCCTCGAGATCCGGCAGCGCCGAGGCCACGC 2204

Qy         626 GCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCT 685
               || | |  |  |  || ||| || ||| | |  |||| | |  |||| | ||  | |
Db        2203 CTTCCGCCTCGCGCAGCCGCGCGCGGGCCTCCATCGCCTCGTCCTCGTAGCGCTTCACCG 2144

Qy         686 CGTCGCGCGAGATGTCGAGCT--CGATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTC 743
               ||   ||||     |||||||  |   ||| |  ||   |  |    |||  || | |
Db        2143 CGAGCCGCGCCGATTCGAGCTTGCCGCCCGAGATCGAACGGTCCGCCTCGGCGCGGGCCG 2084

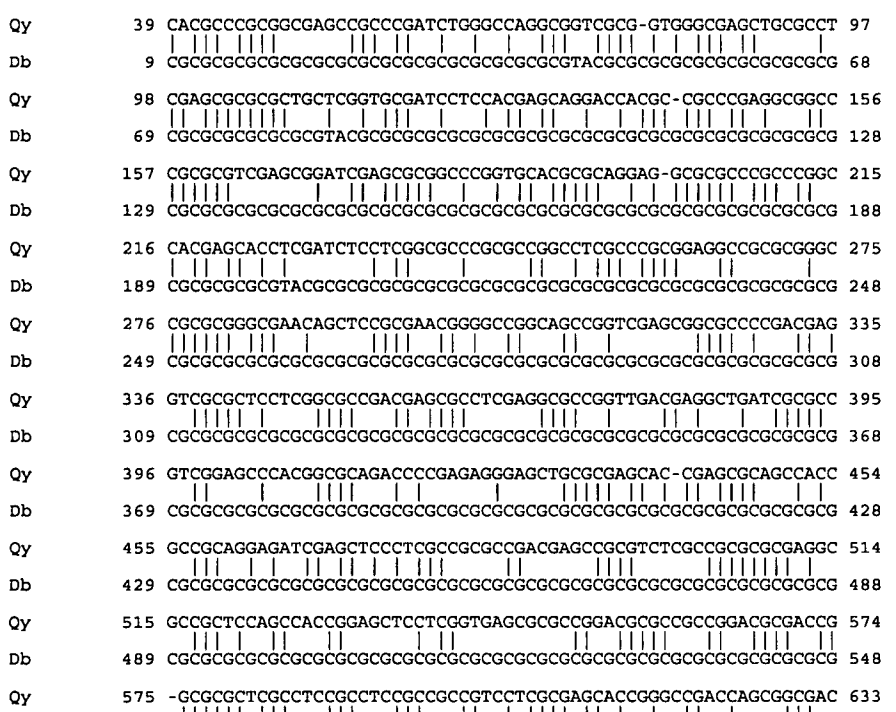
Qy         744 GCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGC 803
               ||  ||    ||| ||    |  |  |  | |||  ||  ||  |  |||     ||
Db        2083 CCGCACGGCTCGCCTCGGTCACGCGCGCATCGGCGGCGCGGCGGGCCTCGCGGGCCTCCT 2024

Qy         804 GCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTC 863
               |  | | | | | | ||| |  | || | || |||| | |  |    |  || | | |
Db        2023 GATCGGCGCGGGTGAGCTGCGCGAGCCGGGCCTGCAGCGCCTCGTGCGCCTGCCTTGCCC 1964

Qy         864 CCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCG 897
               ||| | ||  || |||  |  |     |||| ||
Db        1963 CCTCGGCGCGGGCCGCCACCTCCTCGAGATCGCG 1930
```

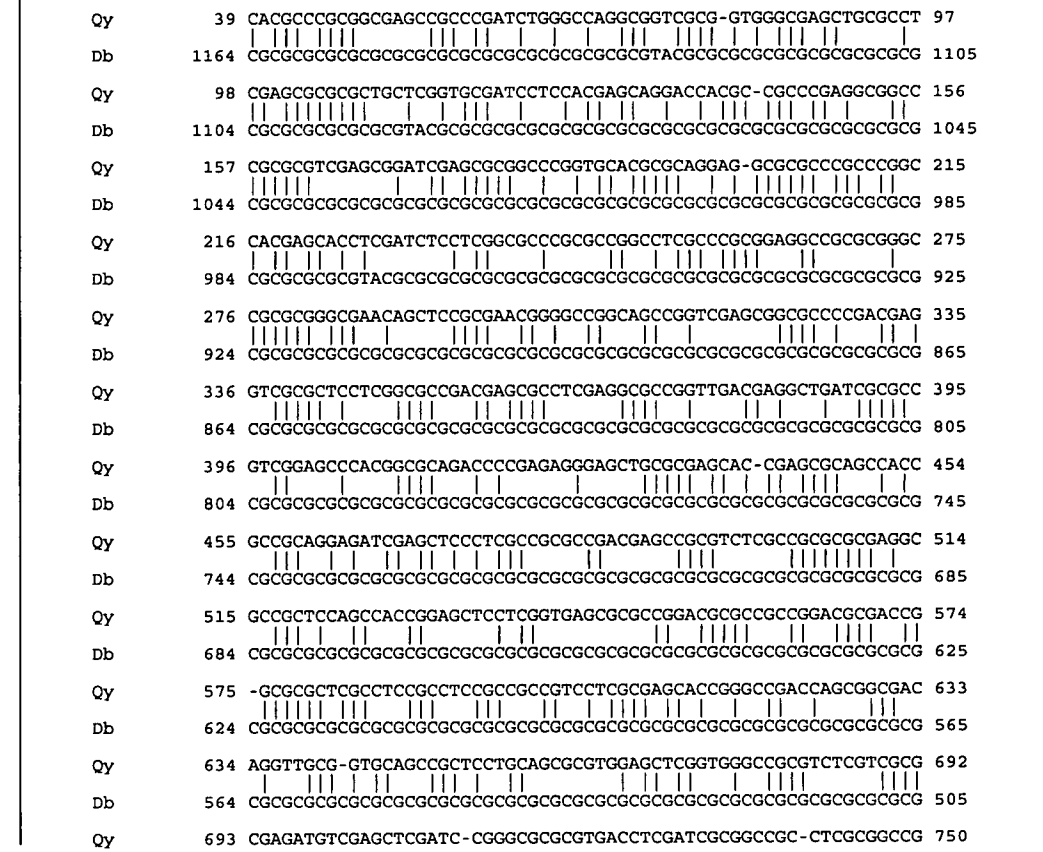
RESULT 4
US-10-363-345A-18585
; Sequence 18585, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18585
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18585
US-10-363-345A-18585

Query Match 10.7%; Score 107.6; DB 8; Length 1172;
Best Local Similarity 51.7%; Pred. No. 1.9e-13;
Matches 504; Conservative 0; Mismatches 459; Indels 11; Gaps 11;



RESULT 5
US-10-363-345A-18586/c
; Sequence 18586, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18586
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18586
US-10-363-345A-18586

Query Match 10.7%; Score 107.6; DB 8; Length 1172;
Best Local Similarity 51.7%; Pred. No. 1.9e-13;
Matches 504; Conservative 0; Mismatches 459; Indels 11; Gaps 11;



```
Db        504 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 445

Qy        751 AGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCC 810

Db        444 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 385

Qy        811 ACCGAGGGGCCGATCTCGCG-GCGCGTGCACAGGCGGGCCGCCTCGCGCTC-CTCCCTGG 868

Db        384 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 325

Qy        869 CCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGA 928

Db        324 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 265

Qy        929 ACGTCGGGAGCGACCACGGGTCGATGGGCGC-CGCGCCCGCCGGCGCGCCCGCGCCCCCG 987

Db        264 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 205

Qy        988 CGCAGGAGCGCGAG 1001

Db        204 CGCGCGCGCGCGCG 191


RESULT 6
US-10-363-345A-18587/c
; Sequence 18587, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18587
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18587
US-10-363-345A-18587

  Query Match             10.7%;  Score 107.6;  DB 8;  Length 1172;
  Best Local Similarity   51.7%;  Pred. No. 1.9e-13;
  Matches  504;  Conservative    0;  Mismatches  459;  Indels   11;  Gaps   11;

Qy         39 CACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCG-GTGGGCGAGCTGCGCCT 97

Db       1102 CGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1043

Qy         98 CGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGC-CGCCCGAGGCGGCC 156

Db       1042 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 983

Qy        157 CGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAG-GCGCGCCCGCCCGGC 215

Db        982 CGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 923

Qy        216 CACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGC 275

Db        922 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863

Qy        276 CGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAG 335

Db        862 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803

Qy        336 GTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCC 395

Db        802 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743

Qy        396 GTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCAC-CGAGCGCAGCCACC 454

Db        742 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683

Qy        455 GCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGC 514

Db        682 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623

Qy        515 GCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCG 574

Db        622 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 563

Qy        575 -GCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGAC 633

Db        562 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 503

Qy        634 AGGTTGCG-GTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCG 692

Db        502 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443

Qy        693 CGAGATGTCGAGCTCGATC-CGGGCGCGCGTGACCTCGATCGCGGCCGC-CTCGCGGCCG 750

Db        442 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383

Qy        751 AGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCC 810

Db        382 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323

Qy        811 ACCGAGGGGCCGATCTCGCG-GCGCGTGCACAGGCGGGCCGCCTCGCGCTC-CTCCCTGG 868

Db        322 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 263

Qy        869 CCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGA 928

Db        262 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 203

Qy        929 ACGTCGGGAGCGACCACGGGTCGATGGGCGC-CGCGCCCGCCGGCGCGCCCGCGCCCCCG 987

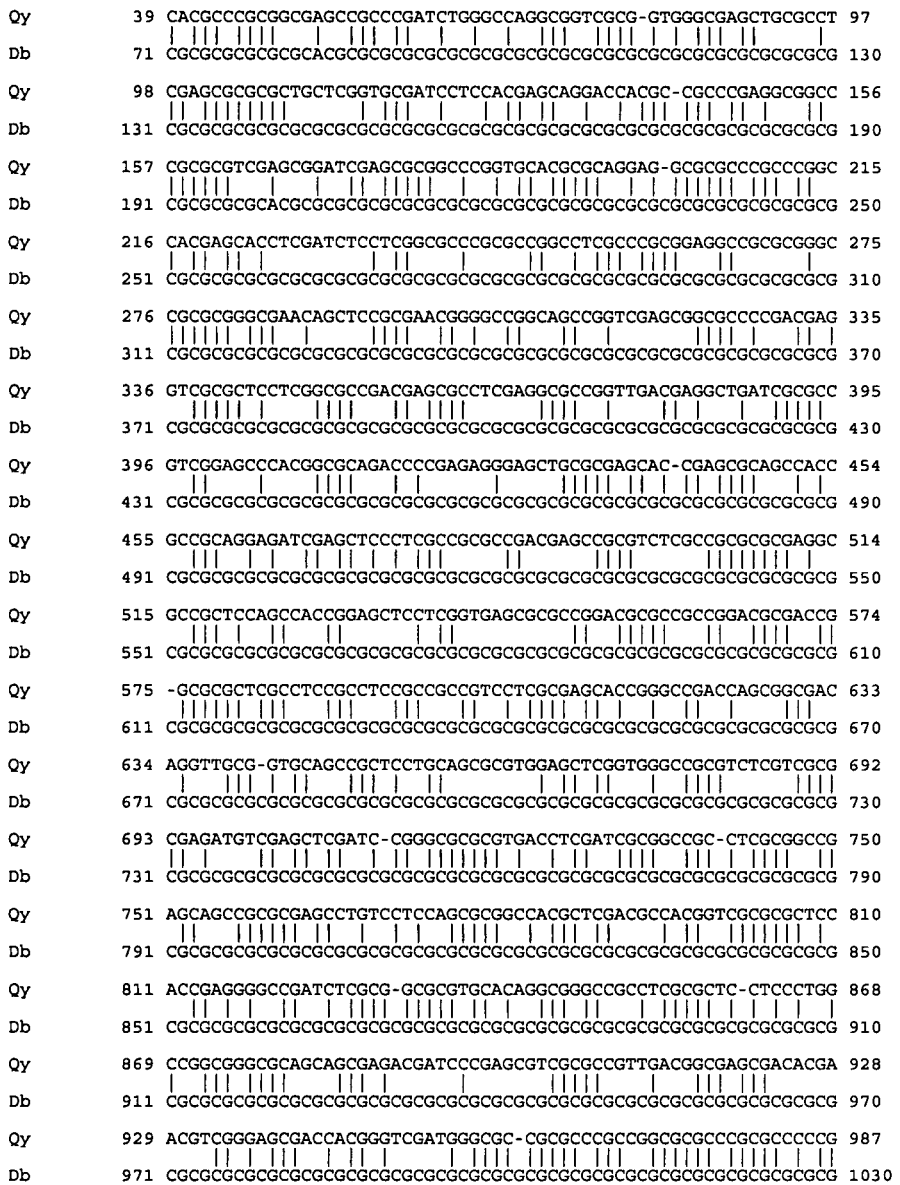
Db        202 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143

Qy        988 CGCAGGAGCGCGAG 1001

Db        142 CGCGCGCGCGCGCG 129


RESULT 7
US-10-363-345A-18588
; Sequence 18588, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18588
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18588
US-10-363-345A-18588

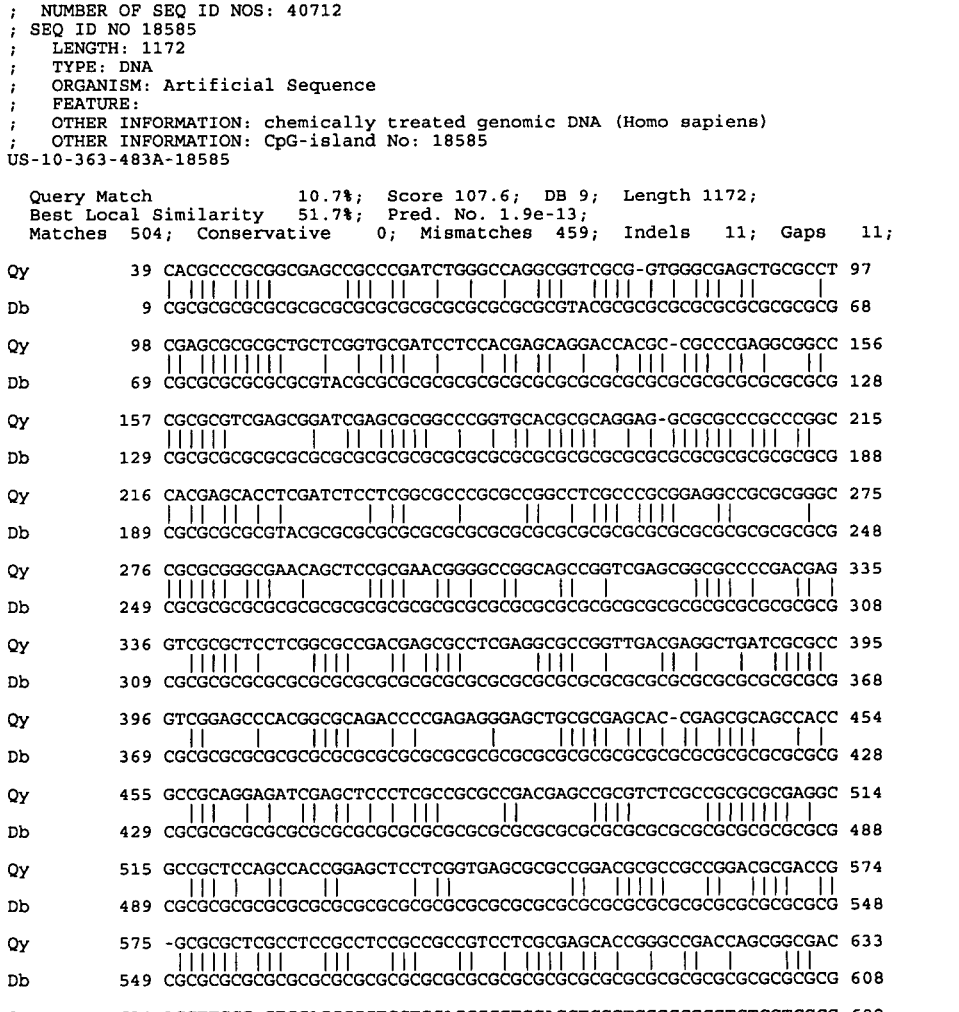
  Query Match             10.7%;  Score 107.6;  DB 8;  Length 1172;
  Best Local Similarity   51.7%;  Pred. No. 1.9e-13;
  Matches  504;  Conservative    0;  Mismatches  459;  Indels   11;  Gaps   11;
```



RESULT 8
US-10-363-483A-18585
; Sequence 18585, Application US/10363483A

; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18585
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18585
US-10-363-483A-18585

Query Match 10.7%; Score 107.6; DB 9; Length 1172;
Best Local Similarity 51.7%; Pred. No. 1.9e-13;
Matches 504; Conservative 0; Mismatches 459; Indels 11; Gaps 11;



```
Db        729 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788

Qy        811 ACCGAGGGGCCGATCTCGCG-GCGCGTGCACAGGCGGGCCGCCTCGCGCTC-CTCCCTGG 868

Db        789 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848

Qy        869 CCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGA 928

Db        849 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908

Qy        929 ACGTCGGGAGCGACCACGGGTCGATGGGCGC-CGCGCCCGCCGGCGCGCCCGCGCCCCCG 987

Db        909 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968

Qy        988 CGCAGGAGCGCGAG 1001

Db        969 CGCGCGCGCGCGCG 982


RESULT 9
US-10-363-483A-18586/c
; Sequence 18586, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18586
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18586
US-10-363-483A-18586

  Query Match            10.7%;  Score 107.6;  DB 9;  Length 1172;
  Best Local Similarity   51.7%;  Pred. No. 1.9e-13;
  Matches  504;  Conservative    0;  Mismatches  459;  Indels   11;  Gaps   11;

Qy         39 CACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCG-GTGGGCGAGCTGCGCCT 97

Db       1164 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCG 1105

Qy         98 CGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGC-CGCCCGAGGCGGCC 156

Db       1104 CGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1045

Qy        157 CGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAG-GCGCGCCCGCCCGGC 215

Db       1044 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 985

Qy        216 CACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGC 275

Db        984 CGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 925

Qy        276 CGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAG 335

Db        924 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 865

Qy        336 GTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCC 395

Db        864 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 805

Qy        396 GTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCAC-CGAGCGCAGCCACC 454

Db        804 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 745

Qy        455 GCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGC 514

Db        744 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 685

Qy        515 GCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCG 574

Db        684 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625

Qy        575 -GCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGAC 633

Db        624 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 565

Qy        634 AGGTTGCG-GTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCG 692

Db        564 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 505

Qy        693 CGAGATGTCGAGCTCGATC-CGGGCGCGCGTGACCTCGATCGCGGCCGC-CTCGCGGCCG 750

Db        504 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 445

Qy        751 AGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCC 810

Db        444 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 385

Qy        811 ACCGAGGGGCCGATCTCGCG-GCGCGTGCACAGGCGGGCCGCCTCGCGCTC-CTCCCTGG 868

Db        384 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 325

Qy        869 CCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGA 928

Db        324 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 265

Qy        929 ACGTCGGGAGCGACCACGGGTCGATGGGCGC-CGCGCCCGCCGGCGCGCCCGCGCCCCCG 987

Db        264 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 205

Qy        988 CGCAGGAGCGCGAG 1001

Db        204 CGCGCGCGCGCGCG 191


RESULT 10
US-10-363-483A-18587/c
; Sequence 18587, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18587
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18587
US-10-363-483A-18587

  Query Match            10.7%;  Score 107.6;  DB 9;  Length 1172;
  Best Local Similarity   51.7%;  Pred. No. 1.9e-13;
  Matches  504;  Conservative    0;  Mismatches  459;  Indels   11;  Gaps   11;

Qy         39 CACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCG-GTGGGCGAGCTGCGCCT 97

Db       1102 CGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1043
```

```
Qy         98 CGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGC-CGCCCGAGGCGGCC 156
              || ||||||||       | |||  |   | || ||  |  | ||| ||| || |   ||
Db       1042 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 983

Qy        157 CGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAG-GCGCGCCCGCCCGGC 215
              ||||||   |   |  || |||||  |  | || |||||  | | |||||| ||| ||
Db        982 CGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 923

Qy        216 CACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGC 275
              | || || |       | ||    |      ||  | ||| ||||   ||      |
Db        922 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863

Qy        276 CGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAG 335
              |||||| |||  |    ||||  || |  ||   ||  |       |||| |   || |
Db        862 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803

Qy        336 GTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCC 395
                |||||  |    ||||   || ||||      ||||  |    || |   |  |||||
Db        802 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743

Qy        396 GTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCAC-CGAGCGCAGCCACC 454
                ||    |    ||||   | |     |     |||||  || | || ||||   | |
Db        742 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683

Qy        455 GCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGC 514
                |||  | |  || || | | |||       ||       |||      ||||||| |
Db        682 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623

Qy        515 GCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCG 574
                |||  |   ||    ||     | ||             ||  |||||   ||  ||||  ||
Db        622 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 563

Qy        575 -GCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGAC 633
                |||||| |||   |||   |||   ||  | |||| || |   |  ||  |    |||
Db        562 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 503

Qy        634 AGGTTGCG-GTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCG 692
                |   ||| | ||   ||| |  ||        | || ||   |  ||||      ||||
Db        502 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443

Qy        693 CGAGATGTCGAGCTCGATC-CGGGCGCGCGTGACCTCGATCGCGGCCGC-CTCGCGGCCG 750
              || |    || || ||  | || |||||||  |  | ||  ||||  ||| | ||||  ||
Db        442 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383

Qy        751 AGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCC 810
                ||   |||||| ||  |  | |  |||||  | ||| ||    | ||  |||||||  |
Db        382 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323

Qy        811 ACCGAGGGGCCGATCTCGCG-GCGCGTGCACAGGCGGGCCGCCTCGCGCTC-CTCCCTGG 868
                || | |  ||  | |||| ||||| || |  ||| ||    | |||||  | | |  |
Db        322 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 263

Qy        869 CCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGA 928
              |  ||| ||||    ||| |      |      |||||    |   ||| |||
Db        262 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 203

Qy        929 ACGTCGGGAGCGACCACGGGTCGATGGGCGC-CGCGCCCGCCGGCGCGCCCGCGCCCCCG 987
                || | |||  | || |     | |||| ||||| |||  |||||| ||||| | ||
Db        202 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143

Qy        988 CGCAGGAGCGCGAG 1001
              |||  | ||||| |
Db        142 CGCGCGCGCGCGCG 129


RESULT 11
US-10-363-483A-18588
; Sequence 18588, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18588
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18588
US-10-363-483A-18588

  Query Match            10.7%;  Score 107.6;  DB 9;  Length 1172;
  Best Local Similarity   51.7%;  Pred. No. 1.9e-13;
  Matches  504;  Conservative    0;  Mismatches  459;  Indels   11;  Gaps   11;

Qy         39 CACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCG-GTGGGCGAGCTGCGCCT 97
              | ||| ||||   |  ||| ||  |  |  | |||  |||| | | ||| ||      |
Db         71 CGCGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 130

Qy         98 CGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGC-CGCCCGAGGCGGCC 156
              || ||||||||       | |||  |   | || ||  |  | ||| ||| || |   ||
Db        131 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 190

Qy        157 CGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAG-GCGCGCCCGCCCGGC 215
              ||||||   |   |  || |||||  |  | || |||||  | | |||||| ||| ||
Db        191 CGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 250

Qy        216 CACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGC 275
              | || || |       | ||    |      ||  | ||| ||||   ||      |
Db        251 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 310

Qy        276 CGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAG 335
              |||||| |||  |    ||||  || |  ||   ||  |       |||| |   || |
Db        311 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 370

Qy        336 GTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCC 395
                |||||  |    ||||   || ||||      ||||  |    || |   |  |||||
Db        371 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 430

Qy        396 GTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCAC-CGAGCGCAGCCACC 454
                ||    |    ||||   | |     |     |||||  || | || ||||   | |
Db        431 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 490

Qy        455 GCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGC 514
                |||  | |  || || | | |||       ||       |||      ||||||| |
Db        491 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550

Qy        515 GCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCG 574
                |||  |   ||    ||     | ||             ||  |||||   ||  ||||  ||
Db        551 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 610

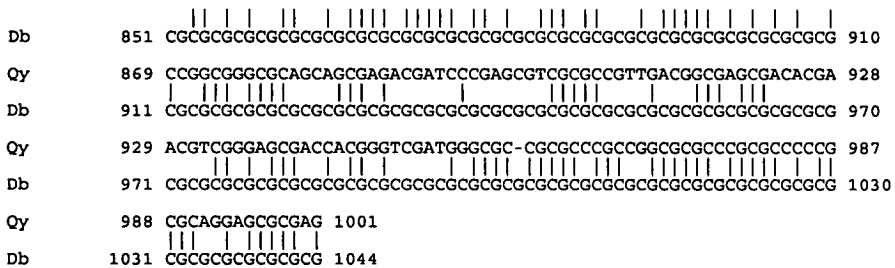
Qy        575 -GCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGAC 633
                |||||| |||   |||   |||   ||  | |||| || |   |  ||  |    |||
Db        611 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 670

Qy        634 AGGTTGCG-GTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCG 692
                |   ||| | ||   ||| |  ||        | || ||   |  ||||      ||||
Db        671 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730

Qy        693 CGAGATGTCGAGCTCGATC-CGGGCGCGCGTGACCTCGATCGCGGCCGC-CTCGCGGCCG 750
              || |    || || ||  | || |||||||  |  | ||  ||||  ||| | ||||  ||
Db        731 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790

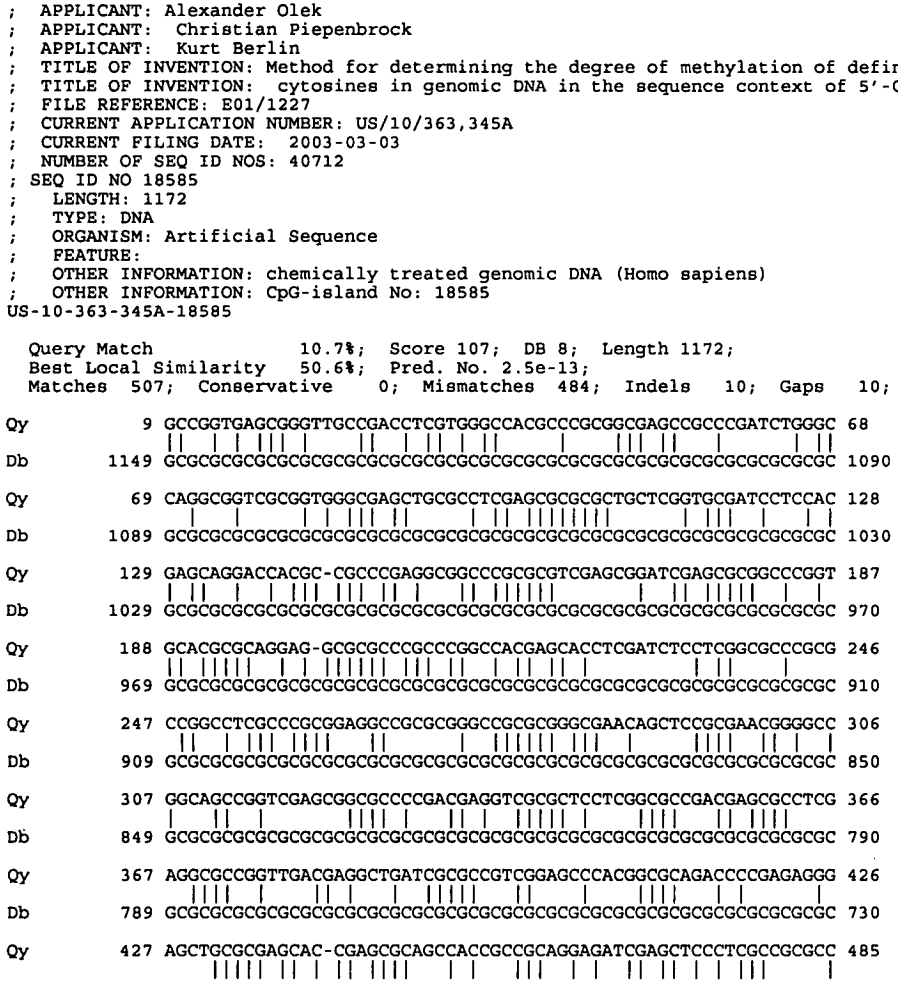
Qy        751 AGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCC 810
                ||   |||||| ||  |  | |  |||||  | ||| ||    | ||  |||||||  |
Db        791 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 850

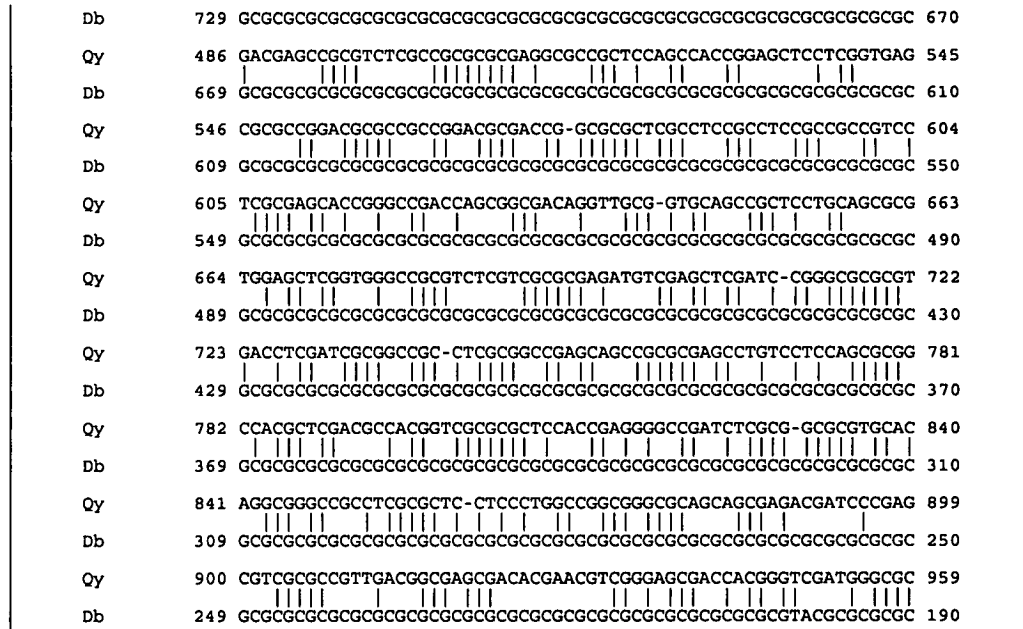
Qy        811 ACCGAGGGGCCGATCTCGCG-GCGCGTGCACAGGCGGGCCGCCTCGCGCTC-CTCCCTGG 868
```



RESULT 12
US-10-363-345A-18585/c
; Sequence 18585, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18585
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18585
US-10-363-345A-18585

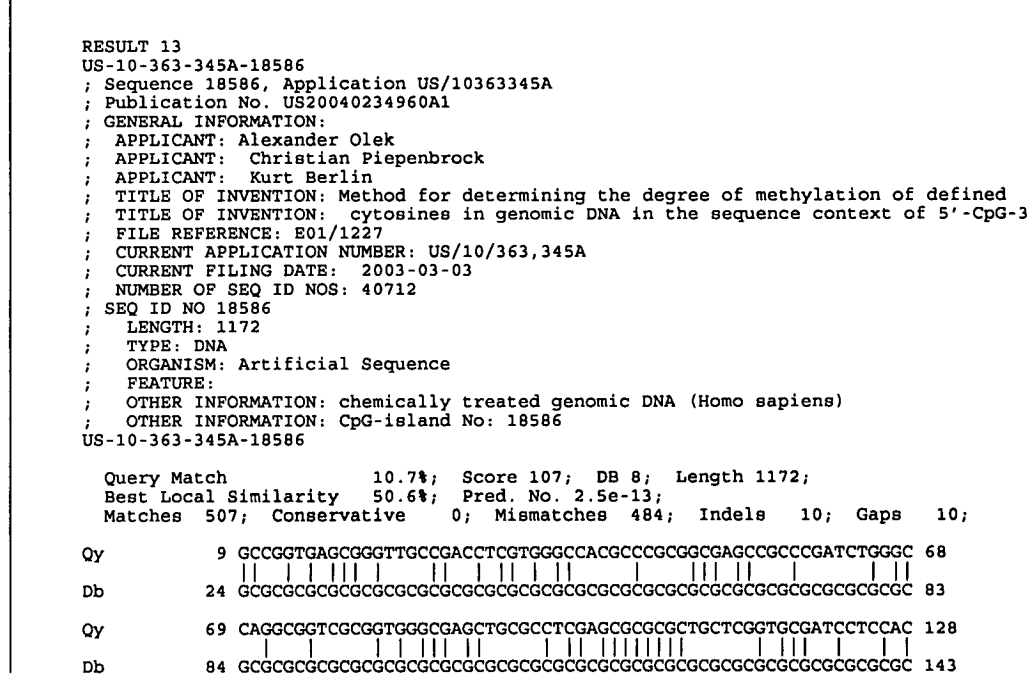
Query Match 10.7%; Score 107; DB 8; Length 1172;
Best Local Similarity 50.6%; Pred. No. 2.5e-13;
Matches 507; Conservative 0; Mismatches 484; Indels 10; Gaps 10;

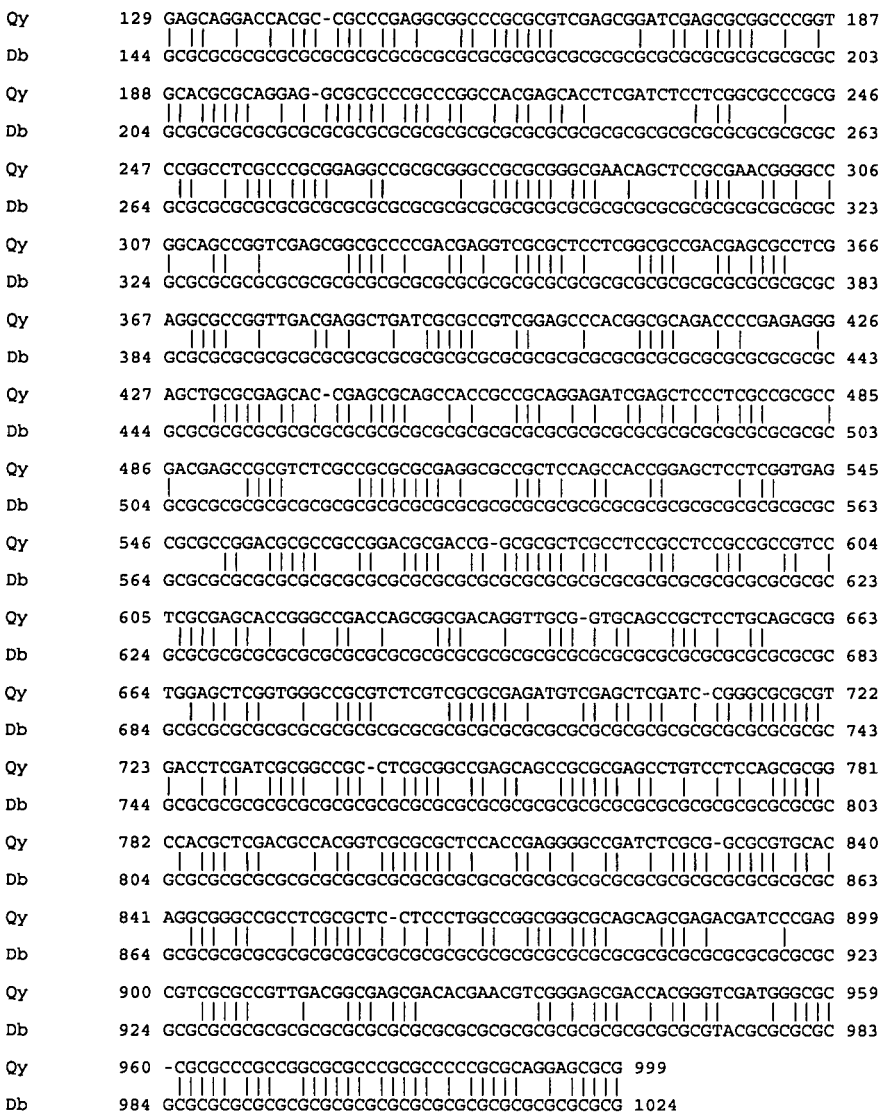




RESULT 13
US-10-363-345A-18586
; Sequence 18586, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18586
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18586
US-10-363-345A-18586

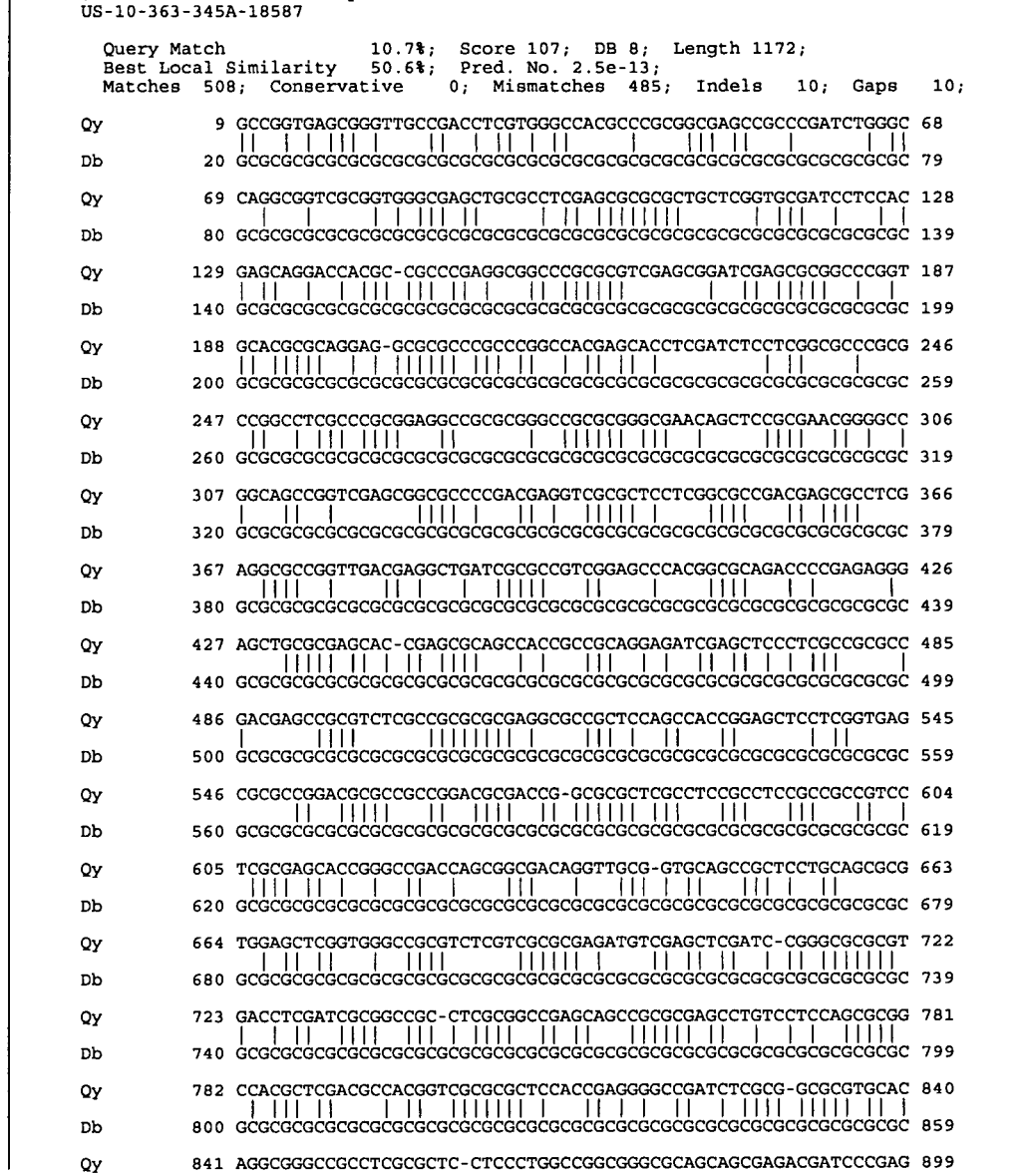
Query Match 10.7%; Score 107; DB 8; Length 1172;
Best Local Similarity 50.6%; Pred. No. 2.5e-13;
Matches 507; Conservative 0; Mismatches 484; Indels 10; Gaps 10;





RESULT 14
US-10-363-345A-18587
; Sequence 18587, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18587
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18587
US-10-363-345A-18587

Query Match 10.7%; Score 107; DB 8; Length 1172;
Best Local Similarity 50.6%; Pred. No. 2.5e-13;
Matches 508; Conservative 0; Mismatches 485; Indels 10; Gaps 10;



```
Db        860 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 919

Qy        900 CGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGC 959

Db        920 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGC 979

Qy        960 -CGCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001

Db        980 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1022


RESULT 15
US-10-363-345A-18588/c
; Sequence 18588, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18588
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18588
US-10-363-345A-18588

  Query Match             10.7%;  Score 107;  DB 8;  Length 1172;
  Best Local Similarity   50.6%;  Pred. No. 2.5e-13;
  Matches  508;  Conservative    0;  Mismatches  485;  Indels   10;  Gaps   10;

Qy          9 GCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGC 68

Db       1153 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1094

Qy         69 CAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCAC 128

Db       1093 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1034

Qy        129 GAGCAGGACCACGC-CGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGT 187

Db       1033 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 974

Qy        188 GCACGCGCAGGAG-GCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCG 246

Db        973 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 914

Qy        247 CCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCC 306

Db        913 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 854

Qy        307 GGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCG 366

Db        853 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 794

Qy        367 AGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGG 426

Db        793 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 734

Qy        427 AGCTGCGCGAGCAC-CGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCC 485

Db        733 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 674

Qy        486 GACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAG 545

Db        673 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 614

Qy        546 CGCGCCGGACGCGCCGCCGGACGCGACCG-GCGCGCTCGCCTCCGCCTCCGCCGCCGTCC 604

Db        613 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 554

Qy        605 TCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCG-GTGCAGCCGCTCCTGCAGCGCG 663

Db        553 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 494

Qy        664 TGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATC-CGGGCGCGCGT 722

Db        493 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 434

Qy        723 GACCTCGATCGCGGCCGC-CTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGG 781

Db        433 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 374

Qy        782 CCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCG-GCGCGTGCAC 840

Db        373 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 314

Qy        841 AGGCGGGCCGCCTCGCGCTC-CTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAG 899

Db        313 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 254

Qy        900 CGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGC 959

Db        253 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGC 194

Qy        960 -CGCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001

Db        193 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151


Search completed: April 27, 2006, 14:39:37
Job time : 1006.57 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 628.575 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_2000_3000
Perfect score: 1001
Sequence: 1 caggcgatgccggtgagcgg.....cccccgcgcaggagcgcgag 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	111.6	11.1	4770	7	US-10-496-351-3	Sequence 3, Appli
	2	111.6	11.1	82746	7	US-10-496-351-56	Sequence 56, Appl
c	3	108.4	10.8	9975	7	US-10-496-351-4	Sequence 4, Appli
c	4	108.2	10.8	11070	14	US-11-075-185-34	Sequence 34, Appl
c	5	108.2	10.8	78869	14	US-11-075-185-1	Sequence 1, Appli
c	6	104.4	10.4	14055	7	US-10-496-351-1	Sequence 1, Appli
c	7	104.4	10.4	88421	14	US-11-205-109-1	Sequence 1, Appli
c	8	104	10.4	9222	7	US-10-496-351-2	Sequence 2, Appli
c	9	103.8	10.4	3354	7	US-10-496-351-12	Sequence 12, Appl
	10	102.8	10.3	1164	7	US-10-496-351-48	Sequence 48, Appl
	11	102.8	10.3	11905	7	US-10-496-351-57	Sequence 57, Appl
c	12	102.2	10.2	10968	14	US-11-075-185-35	Sequence 35, Appl
c	13	100.2	10.0	1554	7	US-10-496-351-52	Sequence 52, Appl
c	14	100	10.0	2742	7	US-10-496-351-22	Sequence 22, Appl
c	15	98.6	9.9	1251	7	US-10-496-351-9	Sequence 9, Appli
c	16	96.4	9.6	5679	14	US-11-075-185-36	Sequence 36, Appl
	17	95.6	9.6	4770	7	US-10-496-351-3	Sequence 3, Appli
c	18	95.6	9.6	82746	7	US-10-496-351-56	Sequence 56, Appl
	19	94.2	9.4	3354	7	US-10-496-351-12	Sequence 12, Appl
c	20	93.4	9.3	2247	7	US-10-496-351-8	Sequence 8, Appli
	21	93.4	9.3	11070	14	US-11-075-185-34	Sequence 34, Appl
	22	93.4	9.3	78869	14	US-11-075-185-1	Sequence 1, Appli
	23	92.6	9.3	14055	7	US-10-496-351-1	Sequence 1, Appli
	24	92	9.2	88421	14	US-11-205-109-1	Sequence 1, Appli
c	25	91.8	9.2	8651	8	US-10-432-483-48	Sequence 48, Appl
	26	91.6	9.2	10968	14	US-11-075-185-35	Sequence 35, Appl
c	27	90	9.0	1047	7	US-10-496-351-11	Sequence 11, Appl
c	28	89.4	8.9	113193	7	US-10-478-943D-1	Sequence 1, Appli
c	29	89.2	8.9	1074	7	US-10-496-351-19	Sequence 19, Appl
c	30	87.8	8.8	11905	7	US-10-496-351-57	Sequence 57, Appl
c	31	87.6	8.8	1773	11	US-11-096-568A-22025	Sequence 22025, A
	32	87.6	8.8	113193	7	US-10-478-943D-1	Sequence 1, Appli
c	33	87.2	8.7	1443	7	US-10-496-351-15	Sequence 15, Appl
	34	87.2	8.7	37507	8	US-10-522-037-2	Sequence 2, Appli
c	35	86.8	8.7	1110	8	US-10-432-483-14	Sequence 14, Appl
c	36	86.8	8.7	1506	8	US-10-971-560-9	Sequence 9, Appli
c	37	86.8	8.7	1506	14	US-11-076-427A-1	Sequence 1, Appli
c	38	86.8	8.7	3289	8	US-10-971-560-8	Sequence 8, Appli
c	39	86.8	8.7	4158	8	US-10-971-560-3	Sequence 3, Appli
c	40	86.8	8.7	6458	8	US-10-971-560-1	Sequence 1, Appli
	41	86	8.6	9975	7	US-10-496-351-4	Sequence 4, Appli
c	42	85.4	8.5	1209	7	US-10-496-351-29	Sequence 29, Appl
	43	85.2	8.5	8651	8	US-10-432-483-48	Sequence 48, Appl
c	44	84.8	8.5	116856	14	US-11-143-980-1	Sequence 1, Appli
	45	84.2	8.4	9222	7	US-10-496-351-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-496-351-3/c
; Sequence 3, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4770
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-3

Query Match 11.1%; Score 111.6; DB 7; Length 4770;
Best Local Similarity 47.4%; Pred. No. 1.9e-11;
Matches 463; Conservative 0; Mismatches 504; Indels 9; Gaps 4;

```
Qy          3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
              || | |||    |    ||  | || | ||| |    | ||  | | || ||   | |
Db       4092 GGTGGTGCTCTCGTCGCGGGCGACGGCGTCGAGCAGCGCGGACACCGCCAGGTCCGGGGG 4033

Qy         63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATC 122
              |    |  ||  |  |||||||||| ||  |||   |  | |  || ||
Db       4032 CATCAGGCGCAGGCCAGCCTGGGCGAGCTGGGCGGCGA---TGTCCGGGGTGGCGCTGGA 3976

Qy        123 CTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGC 182
              |    ||     | ||  | |||||  ||||| |    | | |||   |  || |||||
Db       3975 CATGCCGCCGCCGTCCCAGGCGCCCCAGGCGACGGAGGTGGCGACGCGGCCGTGCGCGCG 3916

Qy        183 CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCC 242
              ||| ||  |  |||   |       ||  ||     ||    | | |      || | ||||
Db       3915 GCGGCGCTGGGCCAGTGCGTCGAGGCGGGCGTTGGCGGCGGCGTAGGCGCACTGGCCGCC 3856
```

```
Qy        243 CGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGG 302
Db       3855 GCTGCCCCACACGCCCGCGTTGGAGGAGAACAGGACGAACGCGTCGAGCGGGCGCCCGCC 3796

Qy        303 GGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGC 362
Db       3795 CAGCAGCTCGTCGAGGTGCTCGGCGCCGAGCGCCTTGGCGGCCTGGACGCGCGCGTGCTC 3736

Qy        363 CTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAG 422
Db       3735 CTCGGGGGTGGTGTCGGCGACGGGGGTCTGGGCGTTGGCGCCCGCCGCGTGCACGACGGC 3676

Qy        423 AGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGC 482
Db       3675 GGTGACCTGCTCGCCCGCGCGCTCCAGGCCGGTCAGCAGGGCGGCGAGCGCGGCGCGGTC 3616

Qy        483 GCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGT 542
Db       3615 GGCGACGTCGCAGG---CGGCGGTGGTGACGCGGGCGCCCAGCGCCTCCAGGTCGGCGGC 3559

Qy        543 GAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGT 602
Db       3558 GAGCGCGGGCGCGCCGGGGGCCGA-GGGGCCGCGGCGGCTGGTGAGGACGAGGTGGTCGG 3500

Qy        603 CCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGC 662
Db       3499 CGCCGAGGGCGGCGAGGCGGCGGGCCACGTGCGCGGCGACGCCGCCGGTTCCGCCGGTGA 3440

Qy        663 GTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGT 722
Db       3439 TCAGGACCGTGCCGCGCGGGGTCCACGGGGCGGCCGGGGCCTGCGGTTCGGCGCGCACGA 3380

Qy        723 GACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGC 782
Db       3379 GCCTGCGCGCGAACACGCCCTGCGCGCGGATGGCGACCTGGTCCTCACCTCCGTCGTCGA 3320

Qy        783 CACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAG 842
Db       3319 CGCCGGGTCCGGCGAGGACGGCGGCGAGCCCGCGGCGCTGGCGCGGGTCGGCGGGGTCGG 3260

Qy        843 GCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGT 902
Db       3259 CGAGGTCGACCAGCCCGCCCCAGCGGCCG--GGGTGCTCCAGCGCGACGACGCGGCCCAG 3202

Qy        903 CGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGC 962
Db       3201 TCCCCAGGTGGCGGCCTGGGCGGGGTCCGGCGGGCCGTCGTGCGGGGAGATCGCGACGGC 3142

Qy        963 GCCCGCCGGCGCGCCC 978
Db       3141 CCCGGTGGTGGCGCAC 3126
```

RESULT 2
US-10-496-351-56
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 82746
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

Query Match 11.1%; Score 111.6; DB 7; Length 82746;
Best Local Similarity 47.4%; Pred. No. 7.8e-12;
Matches 463; Conservative 0; Mismatches 504; Indels 9; Gaps 4;

```
Qy          3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
Db      67888 GGTGGTGCTCTCGTCGCGGGCGACGGCGTCGAGCAGCGCGGACACCGCCAGGTCCGGGGG 67947

Qy         63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATC 122
Db      67948 CATCAGGCGCAGGCCAGCCTGGGCGAGCTGGGCGGCGA---TGTCCGGGGTGGCGCTGGA 68004

Qy        123 CTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGC 182
Db      68005 CATGCCGCCGCCGTCCCAGGCGCCCCAGGCGACGGAGGTGGCGACGCGGCCGTGCGCGCG 68064

Qy        183 CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCC 242
Db      68065 GCGGCGCTGGGCCAGTGCGTCGAGGCGGGCGTTGGCGGCGGCGTAGGCGCACTGGCCGCC 68124

Qy        243 CGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGG 302
Db      68125 GCTGCCCCACACGCCCGCGTTGGAGGAGAACAGGACGAACGCGTCGAGCGGGCGCCCGCC 68184

Qy        303 GGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGC 362
Db      68185 CAGCAGCTCGTCGAGGTGCTCGGCGCCGAGCGCCTTGGCGGCCTGGACGCGCGCGTGCTC 68244

Qy        363 CTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAG 422
Db      68245 CTCGGGGGTGGTGTCGGCGACGGGGGTCTGGGCGTTGGCGCCCGCCGCGTGCACGACGGC 68304

Qy        423 AGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGC 482
Db      68305 GGTGACCTGCTCGCCCGCGCGCTCCAGGCCGGTCAGCAGGGCGGCGAGCGCGGCGCGGTC 68364

Qy        483 GCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGT 542
Db      68365 GGCGACGTCGCAGG---CGGCGGTGGTGACGCGGGCGCCCAGCGCCTCCAGGTCGGCGGC 68421

Qy        543 GAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGT 602
Db      68422 GAGCGCGGGCGCGCCGGGGGCCGA-GGGGCCGCGGCGGCTGGTGAGGACGAGGTGGTCGG 68480

Qy        603 CCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGC 662
Db      68481 CGCCGAGGGCGGCGAGGCGGCGGGCCACGTGCGCGGCGACGCCGCCGGTTCCGCCGGTGA 68540

Qy        663 GTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGT 722
Db      68541 TCAGGACCGTGCCGCGCGGGGTCCACGGGGCGGCCGGGGCCTGCGGTTCGGCGCGCACGA 68600

Qy        723 GACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGC 782
Db      68601 GCCTGCGCGCGAACACGCCCTGCGCGCGGATGGCGACCTGGTCCTCACCTCCGTCGTCGA 68660

Qy        783 CACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAG 842
Db      68661 CGCCGGGTCCGGCGAGGACGGCGGCGAGCCCGCGGCGCTGGCGCGGGTCGGCGGGGTCGG 68720

Qy        843 GCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGT 902
Db      68721 CGAGGTCGACCAGCCCGCCCCAGCGGCCG--GGGTGCTCCAGCGCGACGACGCGGCCCAG 68778

Qy        903 CGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGC 962
Db      68779 TCCCCAGGTGGCGGCCTGGGCGGGGTCCGGCGGGCCGTCGTGCGGGGAGATCGCGACGGC 68838

Qy        963 GCCCGCCGGCGCGCCC 978
Db      68839 CCCGGTGGTGGCGCAC 68854
```

RESULT 3
US-10-496-351-4/c
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 9975
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-4

Query Match 10.8%; Score 108.4; DB 7; Length 9975;
Best Local Similarity 48.2%; Pred. No. 5.5e-11;
Matches 461; Conservative 0; Mismatches 481; Indels 14; Gaps 5;

```
Qy        41 CGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGA 100
             || | |||| | |      |    || || |||| |||||| || |    ||||   ||
Db      4309 CGGCGGCGGTGGCCTTGGGGGCGAGGACCCGGCGCACGCGGTCGGGGGTGAGCGCGGTGA 4250

Qy       101 GCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCG 160
             || ||  |  |||        | ||| | ||| |||  ||  |  || |   ||    |
Db      4249 TCGTGCCGTCGTCGAGCGCGGCGGCCAGGTGCACGACGCCGGTGAGCGGGTGCGC---GG 4193

Qy       161 CGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGA 220
             ||   || ||  ||||||||||||   | |  ||   ||  |    |    ||||   ||
Db      4192 CGGGGACCGCGTCGAGCGCGGCGGCGAGGGCGGCGCGGTCGGCGACGTCGACGGCGCGGA 4133

Qy       221 GCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGC 280
             ||||  ||     |  | |||||  |       || || ||  |  |||||||| || ||||
Db      4132 CCACCACGCGGGCGCCCGCGCCGGTCAGCTCCGCGACCAGGTCGGCCGCGCCGGGCGCGT 4073

Qy       281 GGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGC 340
             || |  | |      || | |  | | || ||     | ||   |||| |||||||| ||
Db      4072 CGGGGCCCCGGCGGCTGAGCAGCAGCAGGTGGCGCGCGCCGCGGGCGGTGACGAGGTGGC 4013

Qy       341 GCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGG 400
             |  |  ||||  || ||||| |  ||  ||  | | |||| ||| | |  |  || |
Db      4012 GGGCGACGGCCGCGCCGAGCACGCCGGTGCCGCCGGTGACCAGGACGGTGCCCTCGGGGT 3953

Qy       401 AGCCCACGG----CGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGC 456
             ||||      ||   | | | |  || || |||| | | | | |||  | | | ||
Db      3952 CCAGGACGGGTGCGGCGCCCTCGGGGGCGGCGCGGCGCAACCTCGGCGCGGTGACCCGGC 3893

Qy       457 CGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGC 516
             |  |  |  | | |    ||  ||     | | || |      |    | |  ||
Db      3892 CCGCGCGCACGGCGAGCTGGGGCTCCGGTCCGGCGGCGAGCGCGGTGATCAGGTCGTCGG 3833

Qy       517 CGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGC 576
             |    ||||  ||| |  |  | |    ||  |||| || ||  |  | |  |  |||
Db      3832 GGGCGGAGCCGTCGGTGTCCACCAGGA--CGATCCGGCCGGGCTCCTCGGCCTGGGCGGA 3775

Qy       577 GCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGG 636
             |||  ||   |||  | | | || | | |||||     || |||| || |   |     |
Db      3774 GCGGACGAGGCCG--TGCACGGCGGACGTCGCGGCGGTCGCGCCGTCCGGGCCCTCGCCG 3717

Qy       637 TTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAG 696
             |  | | | |    | ||| || || |  |    |||   |  |   | | |
Db      3716 CCCCTGGTGACCACGAGCAGCACCGAGTCGGCGTGCCGGGGATCGGCGAGGAACCCCTGC 3657

Qy       697 ATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGC 756
             |   ||     | |  |    || | |||||   | ||||||   |  | || |  ||
Db      3656 ACCACGGCGAGGGTGGAGCACACCGCGTCCTCGGCGGAGGCCGCGGGGACGTCGTGGCGC 3597
```

```
Qy       757 CGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAG 816
             || |  |||||  | |   |||||||  | ||   ||    ||  ||  |  ||||
Db      3596 AGCACCGGCCTGGGCGCGGTCGCGGCCGGGGTCAGGGCGGGCGTTTCGGTCCACACCACC 3537

Qy       817 GGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGG 876
               |   | |   ||  |||  | |  |  || |  ||||  | ||   |   |  | ||
Db      3536 TCGAACAGCGACCGCGGCGCCCGCGTGGCGGTCAGGTCGCCCGCCGGGCGCAGCACCAGG 3477

Qy       877 CGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGG 936
                 ||  | |||   | || || |  || ||  | |||      | ||    |  |  |
Db      3476 GAGTCCACGGTGAC---CACGGGCCTGCCGTCGCCGTCGGACGCGGTCAGCGCCAGCTCG 3420

Qy       937 AGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAG 992
             || || ||||  | |  ||  | ||| |  || |    || ||    || ||||
Db      3419 CCCGCCCTCGGGGTGGTCAGCCGCACGCGCAGCGCCCGCGCCCCGGTGGCGTGCAG 3364
```

RESULT 4
US-11-075-185-34/c
; Sequence 34, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34
; LENGTH: 11070
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-34

Query Match 10.8%; Score 108.2; DB 14; Length 11070;
Best Local Similarity 47.4%; Pred. No. 5.8e-11;
Matches 464; Conservative 0; Mismatches 498; Indels 17; Gaps 4;

```
Qy        38 CCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCT 97
             || |  |  ||  |  |  ||| |    |  || |  ||||||  |  ||||  ||
Db      8917 CCGCTTCGCCGCGGTACGTCCCCAGGACGTTCAAGACATCGCGGAAGTTGAGCCCCGTAG 8858

Qy        98 CGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCC 157
             || |  |    ||  |    |  ||  | ||||| ||  ||   |  |  ||| ||
Db      8857 CGCGGACCGCGAGCCGGACCTGCCCCGGCCCGAGCGGGCGCAGCGCATCGGGAGCGTCCG 8798

Qy       158 GCGCGTCGAGCGGATCGAGCGCGGC---CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGG 214
             || |||||  |||| || | |    |  | | |  ||||  |  | |  ||| |
Db      8797 CGAGGTGGAGCGCGTCGAACGAGCCTTTCGTCCGGATGGACAGGCGAAGAGAGCGCGCCC 8738

Qy       215 CCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGG 274
             |  || ||  | | |  |||  || ||  | |||||||| || |    |  | ||| || |
Db      8737 CGTCGGGCGGCGCCAGCTCTCCGTCGGTCTCGCCGGCCGCGTCGAGGCGCGGCGCCCGCG 8678

Qy       275 CCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGA 334
             || |||  |||  |||| |        ||  ||      |     ||||| |   |  | ||
Db      8677 CCCCGCCCGCGCGCAGCGCGAGCTCTGGCTCCTCGGCCGCCGTGAGCGCCCGGCGCAAGA 8618

Qy       335 GGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGC 394
             |  |    |||   || |||      |  |||||    ||      |  |  || ||
Db      8617 GCGCCATGTCCGGCGCCCCGGGCCCCAGATCGAGCATGCGAACCGTG-CGATCCGCGTGC 8559

Qy       395 CGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACC 454
             |  ||     ||| ||   ||   ||  || ||| || |     |||  |  ||
Db      8558 TCGACGCGCGTCGCGCGGAGCAGCCCCCAGACCGCAGCGGGACCCAGCGCCGCCACGCCC 8499
```

```
Qy       455 GCCGCAGGAGATCGAGCTCCCTCGCCGCGC----CGACGAGCCGCGTCTCGCCGCGCGCG 510
Db      8498 TCGTCCGGCGCGACCGCCACCGCGCCGCGCGTCACGACGACGAGCTCCGTCGCCGCCAGC 8439

Qy       511 AGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCG 570
Db      8438 TGCGGCGCCGCCAGCCACCCCTGCACCAGGGACAGCGCGTCGCGCGCCGCGGCGTGCACG 8379

Qy       571 ACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGC 630
Db      8378 GCGTCGACGACCGCGCCCGCGGCGTCCGCCGTGAGGTCCACCACCAGGCGGCGTGGCGCC 8319

Qy       631 GACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCG 690
Db      8318 GCCGCGCCGTCCGGCAGCGACGCGAGCAGCTCGGGCAGCCCCGCCACCGCGCGTGCACCG 8259

Qy       691 CGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGC---------CGCC 741
Db      8258 AGCGCCGCCGCCAGCGCGCCGTCGCCGCCGACCACGACGTGCTCCTCCGGCACAAACGCC 8199

Qy       742 TCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTC 801
Db      8198 GTGTCGTCCAGCGCCACGTCCGTCCACGCGACGCGGTACAGATCGCGCTCCGCCGCGCCC 8139

Qy       802 GCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCC 861
Db      8138 GCGGCGGCGGCCGCGCGCACCCGCTCGGCCGTCGTCCACCGCATCTGGAGCGCGCCCACC 8079

Qy       862 TCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGC 921
Db      8078 CGCGCGACCGGGCGGCCCTGCCCGTCGGCCAGCTGGATCGACGCGGTGATCGCGTCGCCA 8019

Qy       922 GACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCG 981
Db      8018 TCCCCACCGGCCTCCACCTCGGCGCGCACCCGCAGCGCCGCCGTCCCACGCGCCTCGAGC 7959

Qy       982 CCCCCGCGCAGGAGCGCGA 1000
Db      7958 GCCACGTCCGACCACGCGA 7940


RESULT 5
US-11-075-185-1/c
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 78869
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-1

  Query Match             10.8%;  Score 108.2;  DB 14;  Length 78869;
  Best Local Similarity   47.4%;  Pred. No. 3.1e-11;
  Matches  464;  Conservative    0;  Mismatches  498;  Indels   17;  Gaps    4;

Qy        38 CCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCT 97
Db     13118 CCGCTTCGCCGCGGTACGTCCCCAGGACGTTCAAGACATCGCGGAAGTTGAGCCCCGTAG 13059

Qy        98 CGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCC 157
Db     13058 CGCGGACCGCGAGCCGGACCTGCCCCGGCCCGAGCGGGCGCAGCGCATCGGGAGCGTCCG 12999

Qy       158 GCGCGTCGAGCGGATCGAGCGCGGC---CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGG 214
Db     12998 CGAGGTGGAGCGCGTCGAACGAGCCTTTCGTCCGGATGGACAGGCGAAGAGAGCGCGCCC 12939

Qy       215 CCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGG 274
Db     12938 CGTCGGGCGGCGCCAGCTCTCCGTCGGTCTCGCCGGCCGCGTCGAGGCGCGGCGCCCGCG 12879

Qy       275 CCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGA 334
Db     12878 CCCCGCCCCGCGCGCAGCGCGAGCTCTGGCTCCTCGGCCGCCGTGAGCGCCCGGCGCAAGA 12819

Qy       335 GGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGC 394
Db     12818 GCGCCATGTCCGGCGCCCCGGGCCCCAGATCGAGCATGCGAACCGTG-CGATCCGCGTGC 12760

Qy       395 CGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACC 454
Db     12759 TCGACGCGCGTCGCGCGGAGCAGCCCCCAGACCGCAGCGGGACCCAGCGCCGCCACGCCC 12700

Qy       455 GCCGCAGGAGATCGAGCTCCCTCGCCGCGC----CGACGAGCCGCGTCTCGCCGCGCGCG 510
Db     12699 TCGTCCGGCGCGACCGCCACCGCGCCGCGCGTCACGACGACGAGCTCCGTCGCCGCCAGC 12640

Qy       511 AGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCG 570
Db     12639 TGCGGCGCCGCCAGCCACCCCTGCACCAGGGACAGCGCGTCGCGCGCCGCGGCGTGCACG 12580

Qy       571 ACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGC 630
Db     12579 GCGTCGACGACCGCGCCCGCGGCGTCCGCCGTGAGGTCCACCACCAGGCGGCGTGGCGCC 12520

Qy       631 GACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCG 690
Db     12519 GCCGCGCCGTCCGGCAGCGACGCGAGCAGCTCGGGCAGCCCCGCCACCGCGCGTGCACCG 12460

Qy       691 CGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGC---------CGCC 741
Db     12459 AGCGCCGCCGCCAGCGCGCCGTCGCCGCCGACCACGACGTGCTCCTCCGGCACAAACGCC 12400

Qy       742 TCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTC 801
Db     12399 GTGTCGTCCAGCGCCACGTCCGTCCACGCGACGCGGTACAGATCGCGCTCCGCCGCGCCC 12340

Qy       802 GCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCC 861
Db     12339 GCGGCGGCGGCCGCGCGCACCCGCTCGGCCGTCGTCCACCGCATCTGGAGCGCGCCCACC 12280

Qy       862 TCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGC 921
Db     12279 CGCGCGACCGGGCGGCCCTGCCCGTCGGCCAGCTGGATCGACGCGGTGATCGCGTCGCCA 12220

Qy       922 GACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCG 981
Db     12219 TCCCCACCGGCCTCCACCTCGGCGCGCACCCGCAGCGCCGCCGTCCCACGCGCCTCGAGC 12160

Qy       982 CCCCCGCGCAGGAGCGCGA 1000
Db     12159 GCCACGTCCGACCACGCGA 12141


RESULT 6
US-10-496-351-1/c
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
```

```
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             10.4%;  Score 104.4;  DB 7;  Length 14055;
  Best Local Similarity   47.6%;  Pred. No. 2.4e-10;
  Matches  477;  Conservative    0;  Mismatches  506;  Indels   19;  Gaps    5;

Qy          4 GCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATC 63
              |||| ||||   ||  ||  | || |   |     | ||    ||    | |||| |
Db      13466 GCGAGGCCGTCCAGGTGGGCGTTGGCCGCCGCGTAGTTGGCCTGCCCGGCGTGGCCGAGC 13407

Qy         64 TGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCC 123
                | ||  |    |  |      |    |||| | ||  | || | |  |||  | ||
Db      13406 GCGCCCGCGACCGAGGAGAACAGCACGAACGCCGCCAGGTCCCGGTCCCTGGTCAGCTCG 13347

Qy        124 TCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCG---CGCGTCGAGCGGATCGAGCGCG 180
              |||| | || || |  |  || ||  || |  | |   |||||| ||| | ||| | |
Db      13346 TCCAGGTGCCGGGCGCCCTCGGCCTTGGGGCGCAGCACCGCGTCCAGCCGCTCGGGGGTC 13287

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
                |  | || |      | |      | |||||||| | || || | | |   |
Db      13286 AGCGCGGTCAGGGCGCCGTCGTCGAGCACCCCGGCCAGGTGCGCCACCAGCGTCAGCGGC 13227

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              |  |    | | || | ||     ||||||    |  ||      |  |   |  ||
Db      13226 CGGTCCGGCGGCACGGCGGCCAGCGCCGCGACCACGTCGGCCCGGTTCGCGACGTCGCCG 13167

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |  |  || ||  ||||     || | || |  |  ||||| |   |  || ||  |||
Db      13166 CGCACGTGCCGCACGTCGGCGCCCGGCACGTCCGGTCCGCGCCCGCTGCGGCTGAGCAGC 13107

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||  |  ||||  |  |     ||  | |  |   |||  |||  | |||| | |  ||
Db      13106 ACCAGGCAGCGCACGCCGTGCGCGCGCACCAGGTGCTCG--GCCAGCAGCGCGGCCAGCG 13049

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCC 480
              |  ||| | ||  | |  | | |||  ||| |||  |  |   |  |  |  |||  |
Db      13048 CGCCGGAACCGCCGGTGATCAGCGCGGCGCCGCCGGGGTCGAGCACGGCCCCGCCTGCGC 12989

Qy        481 GCGCCGACGAG--CCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCT 538
                | ||  |||  |||    |   ||  || ||||    ||  ||   | | | | | |||
Db      12988 CGGGCGCGGAGACCCGGTGCAGCCCCGGCACGAGCACGCGTCCCGCGCGCAGCACCGCCT 12929

Qy        539 CGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCG 598
              | |  |||||||||  |||  | |||  |   | ||  |||  | |   |   ||  ||
Db      12928 CCGGCAGCGCGCCGGTGGCGAGGGCGGGCAGCAGCGCGGCGGACTCGGGCCGGTCGTCCA 12869

Qy        599 CCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCA 658
                  |  |  | | ||||     ||   |     |       ||| | |  ||  ||
Db      12868 GGTCGGCCAGGACGAACCGGCCTGGGTGCTCGGCGGCGGCCGACCGCACCAGGCCCCACA 12809

Qy        659 GCGCGTGGAGCTCGGTGGG-CCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717
              ||||  |  ||||| | | |||   | || | |     |    |||| |||   || ||
Db      12808 CCGCGGCGCCCTCGGCGCGTCCGGCGCGCGGCCCCACCGCGCCCGAGGTCGTCACGAGCA 12749

Qy        718 CGCGTGACCTCGATC----------GCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCC 766
                || | |  |  |           |   || |||| ||   ||| || ||  |   |
Db      12748 GCCGCGTCGACTCCCAGCGGGGCTCGGTCACCACCTCCTGGAGCAGCGGCAGCAGGTCGC 12689

Qy        767 TGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCT 826
              | | | |   ||| | || | |  | ||   | | | |||||  |       ||  | |
Db      12688 GGACGTGCCCGGCGTCGACCCGCACCACCGGCGGCACCCGCTCGCCGACCTCGCGCAGCG 12629

Qy        827 CGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCG 886
                | ||  |      || || || |  ||   ||||| |   || | ||||||||||
Db      12628 GCACGGGCTCGGCGGCGCCGGGCGGCGGGCCGACCTCCGTCCACGCCTGGCGCAGCAGCA 12569

Qy        887 AGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACG 946
                |   ||  |||    ||  | |||| |   | ||||| | ||  | || |  |  |
Db      12568 CCCCCGGCCGCAGCCGGTCGGGGGTGACCGGCCGGGACACCAGCGAGGCGATCTCCAGCA 12509

Qy        947 GGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCCGC 988
                |  |  |  || ||  |||        ||  | | ||||
Db      12508 GCGGGCGGCCCCGCGGGCAGGCCACGCGCACCCGGACGCCGC 12467


RESULT 7
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Staffa, Alfredo
;  TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
;  FILE REFERENCE: 3002-2US
;  CURRENT APPLICATION NUMBER: US/11/205,109
;  CURRENT FILING DATE:  2005-08-17
;  PRIOR APPLICATION NUMBER: US/09/976,059
;  PRIOR FILING DATE: 2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 88421
;   TYPE: DNA
;   ORGANISM: Actinoplanes sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2077)..(3078)
;   OTHER INFORMATION: ORF 1; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3118)..(4032)
;   OTHER INFORMATION: ORF 2; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4038)..(5048)
;   OTHER INFORMATION: ORF 3; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6665)..(5814)
;   OTHER INFORMATION: ORF 4; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7703)..(6693)
;   OTHER INFORMATION: ORF 5; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9464)..(8130)
;   OTHER INFORMATION: ORF 6; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9691)..(10761)
;   OTHER INFORMATION: ORF 7; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (12751)..(10829)
;   OTHER INFORMATION: ORF 8; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13617)..(12802)
;   OTHER INFORMATION: ORF 9; negative strandedness
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (15203)..(13614)
;   OTHER INFORMATION: ORF 10; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15591)..(15863)
;   OTHER INFORMATION: ORF 11; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15880)..(19035)
;   OTHER INFORMATION: ORF 12; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19032)..(39713)
;   OTHER INFORMATION: ORF 13; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (39713)..(65800)
;   OTHER INFORMATION: ORF 14; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65826)..(66530)
;   OTHER INFORMATION: ORF 15; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (66546)..(67370)
;   OTHER INFORMATION: ORF 16; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (67384)..(70059)
;   OTHER INFORMATION: ORF 17; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70099)..(70662)
;   OTHER INFORMATION: ORF 18; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70659)..(71906)
;   OTHER INFORMATION: ORF 19; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (73439)..(71964)
;   OTHER INFORMATION: ORF 20; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (74216)..(73563)
;   OTHER INFORMATION: ORF 21; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75424)..(74213)
;   OTHER INFORMATION: ORF 22; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75535)..(76464)
;   OTHER INFORMATION: ORF 23; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78110)..(76449)
;   OTHER INFORMATION: ORF 24; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79864)..(78107)
;   OTHER INFORMATION: ORF 25; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81624)..(79861)
;   OTHER INFORMATION: ORF 26; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81909)..(81682)
;   OTHER INFORMATION: ORF 27; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82346)..(82062)
;   OTHER INFORMATION: ORF 28; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82587)..(84446)
;   OTHER INFORMATION: ORF 29; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84481)..(85548)
;   OTHER INFORMATION: ORF 30; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85556)..(86845)
;   OTHER INFORMATION: ORF 31; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87372)..(86803)
;   OTHER INFORMATION: ORF 32; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87494)..(88420)
;   OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

  Query Match             10.4%;  Score 104.4;  DB 14;  Length 88421;
  Best Local Similarity   46.1%;  Pred. No. 1.4e-10;
  Matches  461;  Conservative    0;  Mismatches  531;  Indels    8;  Gaps    3;

Qy          4 GCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATC 63
              | |  |||||| ||   | |||  | | || | |    | |   |  |      |  | ||
Db      46995 GGGTCGCCGGTCAGGTCGGTGCGCAGCACGAGCGTGTTGACGAAGAAGCCGACCAGGTTC 46936

Qy         64 TGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCC 123
              |    |  ||||||  | || ||        |   |        |    |||||||   |
Db      46935 TCCAGCGCCCGGTCGGCGCGGCCGGCGACGACGGTGCCGATCGGGACGTCGGTGCCCGCG 46876

Qy        124 TCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCC 183
              | |   ||   | || | |  ||  |||||||  ||  |   |       ||  |||| |
Db      46875 CCGAGCCGCGACAGCAGGACCGCCACGGCGGCCTGCAGGGTCATGAAGACGGTCGCGCCG 46816

Qy        184 CGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCC 243
              |||   ||   |  |||||  | || |  ||||    ||  |||||||  ||  |  |
Db      46815 TGGTCGGCGGCGACCCGGCGCAGGCTCCGGTGCACGTCGGCCGGGATCTCCGCGCTGTGC 46756

Qy        244 GCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGG 303
              | | |   |  || ||  || ||| |   ||  ||| ||     ||    | | |
Db      46755 AGGTGGCCGCGGTGCGAGGCCGCGGCGGGCCGCGGCCGGTCGGCGGGCAGGTCCAGCTCC 46696

Qy        304 GCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCC 363
              ||||||  | |||||||  || | |    | || ||  || |  || |   | |  ||||
Db      46695 TCCGGCACGCCGTCGAGCCGCTCGCGCCAGTGGGCGATCTGCCGGGAGATCAGGCTCGCC 46636

Qy        364 TCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGA 423
              | ||    |  | | ||   |          |  | |   | ||  |  |  | ||
Db      46635 GGGTCGCTGTCGGAGCCCAGCAGGTCACGCTGCCAGAGGGCGTAGTCGGCGTACTGGACG 46576

Qy        424 GGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCG 483
              || ||   | |   | | |   | ||| ||||||| |     ||    |  |||||  |
Db      46575 GGCAGGGTCTCCGACTCGGGCGCCTGCCCCCGCCGCCGCGCGGCGTAGGCGGTCGCCAGG 46516

Qy        484 CCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTG 543
              ||  |  | |||  |      |  |  |     | | |       | ||| ||     |
Db      46515 TCGCGGGCCAGCGGCCGCATCGACCAGCCGTCGGCCGCGATGTGGTGCAGCACCAGCACG 46456

Qy        544 AGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTC 603
              ||| ||  ||   || ||||||  |||| | || ||     |  ||   | ||   | |
Db      46455 AGCACGTGGGTGCCGTCGCCGGGGGCGAGCAGCACGGCACGCAGCGGGATCTCCCGGGCC 46396

Qy        604 CTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCG 663
                |||   |     || ||      |  |        |  | |||  |   | || ||
Db      46395 AGGTCGAAGGCGTACGCGGCCGCTTCGGCGACCGCCTCCTCGACCCGGCCGGCCGGCACG 46336

Qy        664 TGGAGCTCGGTGGGCCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTG 723
```

```
Db      46335 TCCACGACCGGCAACGCGGGCTCGGCCCGCGCGGGGTCGAG---GATGTGTTGCCACGGC 46279

Qy        724 ACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCC 783
              |  ||   ||| |||           | |||||  | || ||| | || |||   ||  |
Db      46278 TCGCCGTCGGCGACCGTGAACACCGTGCGCAGCACCTCGTGCCGGGCCACCACGTCGCGC 46219

Qy        784 ACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGG 843
              | |   | | || ||   | || || |    | |||   | |||  ||| |  | ||
Db      46218 AGGGCGGCGGTCAGGGCATCCCGGTCGA---GGTCGCCCTCCAGGCGCAGCGCGACCGGG 46162

Qy        844 CGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGC--GAGACGATCCCGAGCG 901
              |     |     |   |   || || ||  ||| | |   || ||   | | |||
Db      46161 ATGTTGTAGGTGGCGCTCGGTCCTTCCAGCTGGGCCAGGAACCACAGCCGGCGCTGCGCG 46102

Qy        902 TCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCG 961
              | |   |  |  | || |||   ||  |  |||     | | | || || | | || |
Db      46101 TAGGAAAGCGGCAGCCGCGCGGGCCGCGCACCGGCCCGCAGCGCCGGCCGGTCGTCGGCA 46042

Qy        962 CGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001
              |  ||| || |||| | |   |  |  ||| || | |||| |
Db      46041 CCGCCGTCGCCGCGGCGGGTCAGCGCCGCGGGGGTGCGTG 46002


RESULT 8
US-10-496-351-2/c
; Sequence 2, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 9222
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-2

  Query Match             10.4%;  Score 104;  DB 7;  Length 9222;
  Best Local Similarity   45.7%;  Pred. No. 3.3e-10;
  Matches  445;  Conservative    0;  Mismatches  520;  Indels    9;  Gaps    2;

Qy          2 AGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGA 61
              || |||  ||   |  | ||   || ||    | ||||    || |    || || |
Db       8854 AGTCGAACCCCAGGTCCCGGAACGCGGCCCGCCCGTCCACCGGGCCGTCCGCGCGGCCCA 8795

Qy         62 TCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGAT 121
                |  |||   |  || |  | |  | |||  |  |   | |  |  | | || | ||
Db       8794 GCACGGCGGCGGCGTGGTCGCGCACCAGCGCCAGCAGCACCCGGTCCCGGTCCGGGCCGG 8735

Qy        122 CCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGG 181
                |   ||| |  | ||    ||  |  ||||| || ||  ||| || |  | | | |||
Db       8734 TCAGGTCGACCGCGTCCGGCTCGGGCTGGGCGGTCCTCGGCTCGGGCAGCACCACCTCGG 8675

Qy        182 CCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCT---CGG 238
                |   || ||  || ||    | || || |||    |||  |   |  ||     | |
Db       8674 GCGTCCGCGCGAACACGACCACCCCCAGCTCGGGGCTCAGCGGCGGCGTGCCCGGTCCCG 8615

Qy        239 CGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGA 298
              ||| |      || | ||  |  |  | || |||||     |||  ||| |    ||   |
Db       8614 CGCACAGCAGGGTCGCGGGCAGGCCGTCCCCGCGGCGGCGGCGCACGAGCGCGGCCAGCA 8555

Qy        299 ACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGA 358
                | ||| | ||  |  |||    ||   | |  ||  | || || |  | ||     ||
Db       8554 CCCCGGCGGCCACCGCGTCCTCGGGGCGCGCCGCGTCGCCGAGCGCGCCCGCCGGTCCGC 8495
```

```
Qy        359 GCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCC 418
              | | |||   ||  |      ||  |    |     |||     ||  |||   | |
Db       8494 ACAGCGCGAACAGCGCGGGCGGCCGGTCGCGGGTCAGCTCGTCCAGCAGCGCGGCCAGCG 8435

Qy        419 CGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCG 478
              | ||        |   | ||  |   |||     ||       |   | | ||   |
Db       8434 CCAGGCCCGCGTCGTCCGGCGCGACGGCGAGCACGACCCAGTCCGGGCGGTGGTCGTCCA 8375

Qy        479 CCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCT 538
              |||| | || |||  || |  |||      ||  | ||| || | | |      | || ||
Db       8374 GCGCGGCCACCAGCTCCGCCCCGGCCCGAGTCGGCGGGCAGCACGACCGGTTCGCGCACCC 8315

Qy        539 CGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCG 598
              ||  |  |||      | |  ||||   |||| ||  | ||  | ||  | ||  | |||
Db       8314 CGCAGGCCGC------CACCACGCCCCGCGCGGCCTCCCCGACCACCAGCACCGGCCCCG 8261

Qy        599 CCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCA 658
                |    |||      || |  |||  || ||  |  |     | |||  |    ||
Db       8260 AGGGGGCGGCGCCGGTCGCGGTGACGCGCCGCAGCCTGGGCGCGAGCACCTCGCCCGCGC 8201

Qy        659 GCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGC 718
              || |     |||||||      |      |||  |||  |     |||  | | |    ||
Db       8200 GCACCACCAGCTCGGGCTCGCCGCAGCCGTGCAGCGCGGGCAGCAGCGCCACCGACGCCG 8141

Qy        719 GCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCG 778
              |   | | ||  |   | |||  | ||  | |      |   ||  |||   |  | |||||
Db       8140 GGTCGTCGTCCACGTCCGCCAGCACGACGTCCGCGCGGTCCGGCAGCGAGCGCACCAGCC 8081

Qy        779 CGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGC 838
              |||||||          || |||| |   |  || || |  || |  ||||    ||  |
Db       8080 CGGCCACCGCGCCGTCCCCGGTGACCACCAGCAGCGGGCCGCTGTGCTCGGCCAGCACGT 8021

Qy        839 ACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGA 898
              |   |  ||||     |||  ||  ||| | ||| | || | |   |   ||| || |
Db       8020 CGGCGACGCTCGCCAGCTCCTCGGCCGAGGCGGTCGGCCCCACCTCCAGCACGGTCGCCG 7961

Qy        899 GCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCG 958
              |    ||||       ||| || ||      |    ||| |   |||   ||  |||||
Db       7960 ACCCGCCGCCCCGAGGCGCGGGCAACCGCTCCAGGCGGACCTCGCACCACGCGCCGGGCG 7901

Qy        959 CCGCGCCCGCCGGC 972
              ||    |||| |||
Db       7900 CCTGCGCCGCGGGC 7887


RESULT 9
US-10-496-351-12/c
; Sequence 12, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 3354
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-12

  Query Match             10.4%;  Score 103.8;  DB 7;  Length 3354;
  Best Local Similarity   46.2%;  Pred. No. 4.9e-10;
  Matches  468;  Conservative    0;  Mismatches  527;  Indels   18;  Gaps    3;

Qy          3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
              || ||  |||||||  |    | || |  | | |  | ||    ||| ||| |  | | |
Db       3213 GGTGAGGCCGGTGACGGCCAGGGCGGCGACCTTGCGCTCGGAGCCGGTGAGGCCGCTGGT 3154
```

```
Qy          63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGC------GCTGCTCGGT 116
Db        3153 GTCGTGCTCGACGGGGTCGGCCGCGTCCGGTTCCAGGCGCGCGCACAGCGGTCGCGCGCC 3094

Qy         117 GCGATCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAG 176
Db        3093 ACACGCCCTGGCCAGGTGCCAGGCCCTGCGCAGCACGGCGCGGGCGCGGCGGTGGTCGCC 3034

Qy         177 CGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTC 236
Db        3033 GAGGGCGTGGTGCGCGGCGCCGAGGCAGGCGAGGGTGGCGGCCAGCTCGTAGCGGTCACC 2974

Qy         237 GGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGC 296
Db        2973 GCAGCGCTCCAGGGCGTCGGCGGCCTCGGCGAGCAGGTGCGGGCGGCGGGTCGGCGGGCT 2914

Qy         297 GAACGGGGCCGGCAGCCGGTCGAGCGGCGCC---------CCGACGAGGTCGCGCTCCTC 347
Db        2913 GAGCGGGGCGAGCGCGCGCAGCGCCAGCGCCCTGGGGCGGCCCGCGTCGGCGCCGGTCGT 2854

Qy         348 GGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCAC 407
Db        2853 GGCGAGCTGGTCGAACACCAGCCGCCGGGCGCGGTCGTGGTTGCCCAGGCGGGTCCACGC 2794

Qy         408 GGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATC 467
Db        2793 CTCGGCGGCCCCGGTGCGCCACGGCGCGAGGCCGGGCAGGTCGAGGCCCCACGAGCGCAG 2734

Qy         468 GAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCA 527
Db        2733 GAGGTCGCCGCACGCCAGGAAGTCCGCGAGCGCCGCGTGGTCCCGCCCGGACGCGAGGTG 2674

Qy         528 CCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTC 587
Db        2673 GTGGTGGCCGCGCGCGTGCAGGTAGTGCGGGGCGGCGGGGCCGTCGAAGAGGGCGTCCGG 2614

Qy         588 CGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAG 647
Db        2613 CACGGGCGCGGTGACGTGCCGCCCGGCCTCCCGGCGCCGCCCGCGCAGGGTGTCGGCGAG 2554

Qy         648 CCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTC 707
Db        2553 GATCAGCGAGCCCAGCGGCAGGCCCACGGACGCGCCCCAGCCGTGCGAGGGGACGGCGTC 2494

Qy         708 GATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCT 767
Db        2493 GAGCGCGGCGCGGGCGTGCCCGGCGGCGCCCGCGAGGTCGCCGAGCCGCAGCGCGA---T 2437

Qy         768 GTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTC 827
Db        2436 CACCGAGGCGGCGGCGGCGAAGGAGGCGACCCAGGTGGGCGCGCGCCGGGCGCGGGCCTC 2377

Qy         828 GCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGA 887
Db        2376 GTCGAGCAGCCGGGCGCAGGCCCGGTCGGCGTCGGCGGCGCGGTCGGCGCGCACCAGGGC 2317

Qy         888 GACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGG 947
Db        2316 TAGCAGGGCGAGCAGCGCGGGTTCCTCGGACCACGGGGAGGTGTGGCTGAGGTGGGCGTG 2257

Qy         948 GTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGA 1000
Db        2256 GCGCAGGTAGTGCGCGGCCTCGGCGGCGGCCCGGCGCTGGTCCCCGAGCGCGA 2204


RESULT 10
US-10-496-351-48
; Sequence 48, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
;   LENGTH: 1164
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-48

  Query Match             10.3%;  Score 102.8;  DB 7;  Length 1164;
  Best Local Similarity   49.6%;  Pred. No. 1e-09;
  Matches  492;  Conservative    0;  Mismatches  477;  Indels   23;  Gaps    8;

Qy          17 GCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGT 76
Db          99 GGGCGGCGCCGCCTACCGGGTGGTCGCCGTGGTCGAGCCGTCCGCGCGGGCCCGCGCCGC 158

Qy          77 CGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGA 136
Db         159 CGTGCTCGGCGACCTCACCGGGGTCGCCGCGCTGGAGCGGGCGGAGGACCTGCGCCGCGA 218

Qy         137 CCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCA 196
Db         219 CGAGGTCGACCTGGCGGTGGTCGCCGTGCCGAACCACCTGCACGCGCCGGTCGCGTCCGG 278

Qy         197 GGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCC--GCGCCGGCCTC 254
Db         279 TCTGCTGCGGCGGGGCGTCCCGGTGTTCCTGGAGAAGCCGCTGTGCCTGGGGTCCGCCGA 338

Qy         255 GCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCG 314
Db         339 GGCCGCCGCGCTCACCGCCGCCGAGCGGGCCGGTGGCGCGGTGCTGCTGGGCGGGAGC-- 396

Qy         315 GTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCG 374
Db         397 GCCGCGCGGCACCGGGCCGACGT-GCGGGCGCTGCGGGAGGTGGCCGCGTCGCTGGGCGC 455

Qy         375 GTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGC 434
Db         456 GCTGCGGCACGTCGACGTGTCCTGGGTGCGGTCCAGGGGCATCCCGGACCGGGGCGGCTG 515

Qy         435 GAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCG 494
Db         516 GTTCACCGACCGCACCCGCTCCGGCGGCGGCGCGCTGGTCGACCTCGGCTGGCACCTGCT 575

Qy         495 CGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGA 554
Db         576 CGACACCGCCCGCGTGCTGCTGCGCGGC-GCCACCGGTGACCCGGAGGGCGCG------- 627

Qy         555 CGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCAC 614
Db         628 -GAGCTGGAGCACGTCACCGGCGCGGTGTCCGCGGACTTCGTCGACGACCACGCGTTCCG 686

Qy         615 CGGGCCGACCAGCGGCGACAGGTTGC-GGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGG 673
Db         687 CGCCGCCTGGCGCGGCGACGGGGCGCAGGCGGACCCGGTGGTCGGCGACGTGGAGGACAC 746

Qy         674 TGGGCCGCGTCTCGTCG--CGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGAT 731
Db         747 CGCCAGGGGATTCCTCGTCACCCGGTCGGGGATCTCGATCGGGGTGCGCGCGGCCTGGGC 806

Qy         732 CGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGA 791
Db         807 CTCGCACCAGGCGCTGGACACCACCGTCGTGCGCGTGGAGGGCAGCGGCGGCACGGCCGA 866

Qy         792 ------CGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCG 845
Db         867 GCTGCGCTGCACCTTCGGCTTCAGCCCGAACCGCGAGGGCCCGTCGCGGCTCCTGCTCAC 926
```

```
Qy        846 GGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGC 905
Db        927 CACCGACGGCCGCACCACCCCGGTCGAGCTGCCCGCCGAGCCGGTGGGCGCCGAGTACGA 986

Qy        906 GCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCC 965
Db        987 CGCCCAGCTCGCGAGCCTGCCCGCCCGCCTGGCCGACCCCGCCGCGCGCGGCGAGGCCGC 1046

Qy        966 CGCCGGCGCGCCCCGCGCCCCCGCGCAGGAGCG 997
Db       1047 GTCCGGGGCGCGCTGGATCGCGGGCGCGATCG 1078


RESULT 11
US-10-496-351-57
; Sequence 57, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
;   LENGTH: 11905
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-57

  Query Match            10.3%;  Score 102.8;  DB 7;  Length 11905;
  Best Local Similarity   49.6%;  Pred. No. 4.9e-10;
  Matches  492;  Conservative    0;  Mismatches  477;  Indels   23;  Gaps    8;

Qy         17 GCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGT 76
Db       6198 GGGCGGCGCCGCCTACCGGGTGGTCGCCGTGGTCGAGCCGTCCGCGCGGGCCCGCGCCGC 6257

Qy         77 CGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGA 136
Db       6258 CGTGCTCGGCGACCTCACCGGGGTCGCCGCGCTGGAGCGGGCGGAGGACCTGCGCCGCGA 6317

Qy        137 CCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCA 196
Db       6318 CGAGGTCGACCTGGCGGTGGTCGCCGTGCCGAACCACCTGCACGCGCCGGTCGCGTCCGG 6377

Qy        197 GGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCC--GCGCCGGCCTC 254
Db       6378 TCTGCTGCGGCGGGGCGTCCCGGTGTTCCTGGAGAAGCCGCTGTGCCTGGGGTCCGCCGA 6437

Qy        255 GCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCG 314
Db       6438 GGCCGCCGCGCTCACCGCCGCCGAGCGGGCCGGTGGCGCGGTGCTGCTGGGCGGGAGC-- 6495

Qy        315 GTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCG 374
Db       6496 GCCGCGCGGCACCGGGCCGACGT-GCGGGCGCTGCGGGAGGTGGCCGCGTCGCTGGGCGC 6554

Qy        375 GTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGC 434
Db       6555 GCTGCGGCACGTCGACGTGTCCTGGGTGCGGTCCAGGGGCATCCCGGACCGGGGCGGCTG 6614

Qy        435 GAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCG 494
Db       6615 GTTCACCGACCGCACCCGCTCCGGCGGCGGCGCGCTGGTCGACCTCGGCTGGCACCTGCT 6674

Qy        495 CGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGA 554
Db       6675 CGACACCGCCCGCGTGCTGCTGCGCGGC-GCCACCGGTGACCCGGAGGGCGCG------- 6726

Qy        555 CGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCAC 614
Db       6727 -GAGCTGGAGCACGTCACCGGCGCGGTGTCCGCGGACTTCGTCGACGACCACGCGTTCCG 6785

Qy        615 CGGGCCGACCAGCGGCGACAGGTTGC-GGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGG 673
Db       6786 CGCCGCCTGGCGCGGCGACGGGGCGCAGGCGGACCCGGTGGTCGGCGACGTGGAGGACAC 6845

Qy        674 TGGGCCGCGTCTCGTCG--CGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGAT 731
Db       6846 CGCCAGGGGATTCCTCGTCACCCGGTCGGGGATCTCGATCGGGGTGCGCGCGGCCTGGGC 6905

Qy        732 CGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGA 791
Db       6906 CTCGCACCAGGCGCTGGACACCACCGTCGTGCGCGTGGAGGGCAGCGGCGGCACGGCCGA 6965

Qy        792 ------CGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCG 845
Db       6966 GCTGCGCTGCACCTTCGGCTTCAGCCCGAACCGCGAGGGCCCGTCGCGGCTCCTGCTCAC 7025

Qy        846 GGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGC 905
Db       7026 CACCGACGGCCGCACCACCCCGGTCGAGCTGCCCGCCGAGCCGGTGGGCGCCGAGTACGA 7085

Qy        906 GCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCC 965
Db       7086 CGCCCAGCTCGCGAGCCTGCCCGCCCGCCTGGCCGACCCCGCCGCGCGCGGCGAGGCCGC 7145

Qy        966 CGCCGGCGCGCCCCGCGCCCCCGCGCAGGAGCG 997
Db       7146 GTCCGGGGCGCGCTGGATCGCGGGCGCGATCG 7177


RESULT 12
US-11-075-185-35/c
; Sequence 35, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
;   LENGTH: 10968
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-35

  Query Match            10.2%;  Score 102.2;  DB 14;  Length 10968;
  Best Local Similarity   47.0%;  Pred. No. 6.4e-10;
  Matches  425;  Conservative    0;  Mismatches  468;  Indels   12;  Gaps    3;

Qy          3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
Db       5292 GGCGCGCGCGCCGAGGCGGTTCCTCAGCTCGACGGCCGTGAGCGAGGTGAGCCCCAGCTC 5233

Qy         63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATC 122
Db       5232 CTTCAGCGGCTTGTCCGCGGGGACCTGCGCCGGACCTGCCAGCTGCAGCACGGCCGCCGC 5173

Qy        123 CTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCG------AGCGGATCGAG 176
```

```
Db       5172 CTCCGCGCGCACCAGCTCGAGCAGCGCGCCCGCCCGCTCGGCCTCGGGCAGCGCCGAGAG 5113

Qy        177 CGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTC 236

Db       5112 CCGCGCGCGCAGCGCCGCCGCTCCGGTCGCGGCGGCCGCAGGCGACGCCGCGCGCCGCCC 5053

Qy        237 GGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGC 296

Db       5052 GGCCGCCGGGCGCACGAGCCCGCGCAGCAGCGCCGGCACCTGCCCGGCGTCGCTCGCCGC 4993

Qy        297 GAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGAC 356

Db       4992 ACGCTGGAGCGCCGACAGGTCGAGCGCCGCCGGCACCAGCGCCGCCTCCGGGCGCGCGAG 4933

Qy        357 GAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGAC 416

Db       4932 CGCCCGGTCCAGCAGCTGCATGCCCTGCGCGAGCGCGATCGGCGCGATGCCCTGCCGTCT 4873

Qy        417 CCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCT 476

Db       4872 GAGGCGTGCCAGGTCGGCGGCGCCGAGGTGCGCCGTCATGCCGAGCCCGGCCTGCTCCCA 4813

Qy        477 CGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTC 536

Db       4812 GAAGCCCCAGGCGAGGCTCATCGCCGGCACGCCGCGCGCGCGGAGCTGCGCCGCGAGCGC 4753

Qy        537 CTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGC 596

Db       4752 GTCCAGGAACGCGTTGGCCGCGGCGTAGTTCGCCTGCCCCGCGCTGCCCAGCGTGCCGGC 4693

Qy        597 CGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTG 656

Db       4692 CGCCGACGAGAACAGCACGAAGGCCGCGAGCCGAT---GGCCTCGCGTCGCCGCGGCCAG 4636

Qy        657 CAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGC 716

Db       4635 GTGCCAGCCGCCGTCGATCTTCGGCCGCAGCACGCGCGCGAGCT---GCTCGGCCGAGAG 4579

Qy        717 GCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAG 776

Db       4578 CTTCGTCACGATGACGTCGTCCAGCACGCCCGCCATGTGCACGACCGCCGTGAGGGGGTG 4519

Qy        777 CGCGGCCACGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGT 836

Db       4518 TGCAGCCGGGATCGCCGCGATGACCGCCCCCAGCGCCGCGCCGTCGGCGACGTCGCACGC 4459

Qy        837 GCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCC 896

Db       4458 CGCGACGTCGACCGTCGCGGCGCCCGCGGCGCGCAGCTCGTCCACCAGCGCCGCGGCGTC 4399

Qy        897 GAGCG 901

Db       4398 CGGCG 4394


RESULT 13
US-10-496-351-52/c
; Sequence 52, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 52
;   LENGTH: 1554
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-52

  Query Match             10.0%;  Score 100.2;  DB 7;  Length 1554;
  Best Local Similarity   48.1%;  Pred. No. 2.7e-09;
  Matches  416;  Conservative    0;  Mismatches  433;  Indels   16;  Gaps    4;

Qy        135 GACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCG 194

Db       1409 GACGCCGCGAGCGCGGCCACCCGGTGCTCCGAGCTGGTCAACCGGCCCGTGTCCGACCCG 1350

Qy        195 CAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTC 254

Db       1349 CAGGCCGC-CGACGCCTGCTGCGCGGGCACCAGCCCGTCGGCGGTGGGCATCAGCTCCTC 1291

Qy        255 GCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCG 314

Db       1290 GCACAGCGGCGCCGCCCCGCACATCCGGGCGACGTGCCACCCGCGCCGGACCACCCGGCG 1231

Qy        315 GTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCG 374

Db       1230 CGCCAGCGCGTGCTCGCCCAGCGCCCGCTGCGCCCCTCCCAGGTCGCGCAGGGTCCGCGC 1171

Qy        375 GTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGC 434

Db       1170 CAGCTCGTACTTGTCGTCCACGGCCTCGAACACCGCCACCGCCTCCCGCAGCAGCCGCGG 1111

Qy        435 GAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCG 494

Db       1110 GCGC------CGCTTCACCGGGCCGGTCGCCGCGAGCAGCCGCAGCGCCTGCCCGCGCGC 1057

Qy        495 CGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGA 554

Db       1056 CCGCCTGCTGTCCGAGCCCGGCCTGCCGAGCTGCTCGTGCGCCAGCGCCCGCGCCCGGTC 997

Qy        555 CGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCAC 614

Db        996 CGGGTCGCCCTGCGCCAGCCACGCCTCCGCCGCCGCGGTCCGCCACGGCACCGGCGCGCA 937

Qy        615 CGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGT 674

Db        936 CCCGCCGCCCAGACCCCACCGGGTCAGCTGCTCCCCGCACAGCAGGAAGTCGCCCAGCGC 877

Qy        675 GGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGC 734

Db        876 CGCGCGCGGCCGTCCCCGCGCCAGGAGGAAGTGCCCGCGCGCGTGCAGGTAGTCCACGCT 817

Qy        735 GGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGC 794

Db        816 GG------TGCGGTGCCCGAACACCCCGTTCGGCACCGGCCGCGCCACCACGGCCGCCGC 763

Qy        795 CACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTC 854

Db        762 CTCGTCGCAGCGGCCCATCCTGGTCAGCGCCAGGATCCGGGTGCTCAGCGGCAGCCCCAC 703

Qy        855 GCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGAC 914

Db        702 CGACACCCCCCACGCGGTGGGGGGCAGCACGGTCAGCGCCTCCCGCGCCAGCTCGT---C 646

Qy        915 GGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGC 974

Db        645 CGCCCTCGGGTGGTCGCCCCGGCGCAGCGCCGACTCGGCCGCCAGCGCCAGCACCTGCGC 586

Qy        975 GCCCGCGCCCCCGCGCAGGAGCGCG 999

Db        585 GCGCGGTACCGGCAGCAGCGGCACG 561


RESULT 14
US-10-496-351-22/c
; Sequence 22, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 2742
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-22

  Query Match             10.0%;  Score 100;  DB 7;  Length 2742;
  Best Local Similarity   46.1%;  Pred. No. 2.4e-09;
  Matches  450;  Conservative    0;  Mismatches  515;  Indels   11;  Gaps    3;

Qy          9 GCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGC 68
              || |  ||||  |  |||| |  | | || | ||| | | | | | ||||  | | ||||
Db       1323 GCGGTCGAGCACGCCGCCGCCGACCTCGGGCGCGCGCCCTGGGTGGCGCCGCACCAGGGC 1264

Qy         69 CAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCAC 128
                | |  |  |  ||||        | | ||||||    || ||| ||    |  | |
Db       1263 GTGCCCGGTGGCGACGGCGGCGGTGACGCCCAGCGCGGCGGGCCCGGCGCCGCGCGGCCC 1204

Qy        129 GAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTG 188
              ||   || |  | ||| |  |||| |   |||   | ||     |   ||| |    |
Db       1203 GTGGGCGAGCCGGTCGCGCACGGCGAGCGCGGCGCGCACCGCCCTGGTGGCGTCGTCCTC 1144

Qy        189 CACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCC 248
               |||   ||  |     |   |  ||| |||| || |        ||| ||||    || |
Db       1143 GCCGCTGCGGGTGGAGCCGAACAGGGCGACGATCAGCGACGCGGCCTTGGCGGTGACGAC 1084

Qy        249 GGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGG 308
              | |  |  ||     | || |   |  | |   || |   |  ||    | || | |
Db       1083 GCCGCCCGCCCGCTCGACCTCCTCGCGCACCGCGGTGGCGACGTCGCTCAGCGCGTCGTC 1024

Qy        309 CAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAG 368
              || |   ||| |||| ||  || | | |||| ||||       |  || || |     ||
Db       1023 CACCTCCTCGGGCGGGGCGGCGGCCACGTCGTGCTCGAACCGGGTGAGGACCATGAGGAC 964

Qy        369 GCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAG 428
              ||     ||    |||  |   |     ||||  ||  || |  |  | |||| | ||
Db        963 GCTGACGTGCTTGAGCTCGGCGGTCGGCGCGGACGCCGGGGGGAGGGGCGCGAGCGCGAC 904

Qy        429 CTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGA-GCTCCCTCGCCGCGCCGA 487
              | ||  || | |     | |   | || |   |   || | ||||   | | | ||
Db        903 GGGGGCAGGCGCGGGCGCCGGGGGCGCCTGCCGCTGCGGGACGATCCCGGCCGGGGGCGC 844

Qy        488 CGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCG 547
              || | |  |||||||  || | | |||        | |  |  | || |||   |||
Db        843 CGCGACCGGGGGCGCCGCGACCGGGCCACCG-------CCCTCGGTCGCCCCGGACCGCG 791

Qy        548 CGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCG 607
              || ||| |||  |  ||  |  | ||||     |    |  | |||||          |||
Db        790 CGGCGGGCGCCGCCTCGCGCCGGGCCGGGAGCCGGTGCGGCACCTCCTGGTTGAGGATCG 731

Qy        608 CGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGA 667
              |  || || |  ||  |||          |  |  || |  |||| | || |       |
Db        730 CCCGCTCCAGCTCGCGCAGCTCGCGGCCGGGGTCGAGCCCCAGCTCGTCCACCAGCCTGG 671

Qy        668 GCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGC---TCGATCCGGGCGCGCGTGA 724
              |  |||| ||||   | || || ||| |  | |  ||    ||   ||  | |||
Db        670 CCCTGGTGCGCCGGTACACGCCGAGCGCGTCGGCCTGCCTGCCGCAGTGGTAGAGCGCCC 611

Qy        725 CCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCA 784
              |  || |  | |   |  || || ||||| || ||  |  | ||  | || |   |
Db        610 GCATGAGCTGGCCGCACAGCCGCTCGCGCAGCGGCTCGGCCTCGACCCACCGCTCCAGCT 551

Qy        785 CGCTCGACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGC 844
                ||      |  | ||     |||  |   || |    |   ||||  ||       |||
Db        550 CGTAGATCATCTCGCCGTAGCGGCCGGACGCGAACTCCGCCTCGACGCAGTCCTCCAGCG 491

Qy        845 GGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCG 904
              ||||  |  |   |   |  ||||| | |     ||    ||||  |         |
Db        490 CGGCCAGCCGGGTGTTGCGCACGGCCGCCAGCTCGGGCCAGTCGACGCCCTGCTCGACCA 431

Qy        905 CGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGC 964
              | ||   |      ||       ||    || ||      |   | || | || ||||
Db        430 GGTCGGCCAGCAGCGGCCCGCGCCACAGGCCGATCGCGTCGCGCAGGGTGCGGGCGGCGC 371

Qy        965 CCGCCGGCGCGCCCGC 980
              |  |   | ||||||
Db        370 GCTGCCAGGAGCCCGC 355


RESULT 15
US-10-496-351-9/c
; Sequence 9, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 1251
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-9

  Query Match              9.9%;  Score 98.6;  DB 7;  Length 1251;
  Best Local Similarity   45.7%;  Pred. No. 5.4e-09;
  Matches  387;  Conservative    0;  Mismatches  454;  Indels    6;  Gaps    1;

Qy        159 CGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCAC 218
              |||  |   ||    | |||||   |  | || |     ||   |   ||   |  |||
Db       1179 CGCCGCCGTCGTCACCAGCGCCATCACGACCAGGACGGTGAACAGGCGCCCGTCCAGCAC 1120

Qy        219 GAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGC 278
              ||  ||| | | || || |   | | |   ||||                |||     |
Db       1119 CCCCAGCTCCAGCCCCACGCCCAGCACCACCAGCTCGGTGAGCCCCCTGGCGTTCATCAG 1060

Qy        279 GCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTC 338
              | ||| |     |||||     | ||| |||||||    |||||  |   | ||| |     |
Db       1059 CACGCCGACCTCGGCCGCGTCGCGCGCGGGCAGCCCCGCGAGCAGACCGCCGCCCGCCGC 1000

Qy        339 GCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTC 398
              ||   ||     |||||||  |||  ||| | ||      |                |||||||
Db        999 GCCGACCAGCTTGCCGACGCACGCGGCGACGAGCACCAGCAGCGCCTCCGCCGCGCCGGA 940

Qy        399 GGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCG 458
              |||| | | ||| |||   | |   |||   || |   ||  | |  | | | |  | |
Db        939 CCAGCCGAGGCCGCCGACGTCCACGGACAGCCCGGCCACGACGAAGAACACCGGCAGCAG 880

Qy        459 CAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCG 518
              ||||     |||| | |   ||  | | || ||   || | || |   |  |||
Db        879 CAGGGAGCTGAGCCCGCCGACCCGGTCCACCAGGGTCGGCGCGGCCTCCCGGAGCGCGTC 820

Qy        519 CTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGC 578
              ||   ||         | |      || ||| ||| || |  ||        |||  | ||
Db        819 CTTGGGCACAACCGCCCCGAACAGGAACGCCCCGAACACCGCGTGCAGGCCGATGGCCGT 760

Qy        579 GCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTT 638
              | | ||   |||   |  |  | || |  ||  ||     ||| ||||  ||
Db        759 GGTGGCGAGCGCGAACAGCAGCAGCCCCTGGACGACGACCGCGAACAGCACCGCGTCCGC 700
```

```
Qy        639 GCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGAT 698
              |||  ||   |   ||  |  |  |   | |        | ||| |  | || |    |
Db        699 GCGCGGCGAGCCGCCCACCGTCCGGGCAACCCACCCCACCAGCGGCCGGACGAGGAGCAC 640

Qy        699 GTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTC------GCGGCCGAG 752
              | | |  || | | ||   |||      | | || |||| || |       | ||||| |
Db        639 CGCCACCAGGACCAGCGCCGCCGTCCAGGCCAGCGTGGCCACCGCCGACCAGGGGCCGCG 580

Qy        753 CAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCCAC 812
              |     | ||| || | |  | |  |||  | ||| | |||||||     ||  | || |
Db        579 CGCGGTCACGATCCCGGCGACGACGGCGAGCGCGCACCACGCCACCACGTCGTTCACCGC 520

Qy        813 CGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGG 872
              | |    | | | ||     |  | || ||  |  |  |   |||  |||
Db        519 GGCGCTCGCCAGCGCGACCACCCCGGACCGGGTGCGCTGCATCCGCCGCTCGGCCAGGAT 460

Qy        873 CGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAACGT 932
              | |   |||||  | ||       || || |  | ||    ||| || |  | ||| ||
Db        459 CCGCGCCAGCACGGGGAAAGCGGTGATCGACAGGGCGGCTGCGGTGAACAGCGCGAGCGC 400

Qy        933 CGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAG 992
                  |  |||    |   |||||||| |  || || ||     |  |    |||
Db        399 GTCACCGCCCTGCGGCCACAGCAGCGCCGCGACGCCCAGCCCGAGGCAGAACGGCAGCAC 340

Qy        993 GAGCGCG 999
              || || |
Db        339 GACCGAG 333
```

Search completed: April 27, 2006, 15:52:57
Job time : 630.575 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 4961.09 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_2000_3000
Perfect score: 1001
Sequence: 1 caggcgatgccggtgagcgg.....cccccgcgcaggagcgcgag 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		152	15.2	1885	10	AG363563	AG363563 Mus muscu
2	c	150.8	15.1	1569	10	AG341503	AG341503 Mus muscu
3		149.6	14.9	1674	10	CL078342	CL078342 CH216-149
4		149.4	14.9	1798	10	AG171124	AG171124 Pan trogl
5	c	149	14.9	1878	10	AG393083	AG393083 Mus muscu
6		147	14.7	1299	10	AG039481	AG039481 Pan trogl
7		146.8	14.7	1406	10	CG756569	CG756569 P051-4-B0
8	c	145.2	14.5	1421	9	BZ569488	BZ569488 pacs2-164
9		145	14.5	1421	9	BZ569488	BZ569488 pacs2-164
10		144.8	14.5	2332	10	AG363333	AG363333 Mus muscu
11	c	143.4	14.3	1280	10	CL470317	CL470317 SAIL_140_
12	c	142.4	14.2	1462	10	AG441877	AG441877 Mus muscu
13		142	14.2	1780	10	AG448243	AG448243 Mus muscu
14	c	141.8	14.2	2243	10	AG381986	AG381986 Mus muscu
15	c	141.6	14.1	1531	10	AG430426	AG430426 Mus muscu
16	c	141.6	14.1	1598	10	AG030579	AG030579 Pan trogl
17		141	14.1	1131	10	AG042920	AG042920 Pan trogl
18	c	140.4	14.0	1821	10	CL090560	CL090560 ISB1-17N1
19	c	139.8	14.0	1780	10	AG448243	AG448243 Mus muscu
20		139.4	13.9	1878	10	AG393083	AG393083 Mus muscu
21	c	139.4	13.9	2332	10	AG363333	AG363333 Mus muscu
22	c	139	13.9	1452	10	AG032979	AG032979 Pan trogl
23		139	13.9	1542	10	AG032943	AG032943 Pan trogl
24		139	13.9	1981	10	CL515231	CL515231 SAIL_898_
25	c	138.8	13.9	1438	9	BZ557931	BZ557931 pacs1-60_
26		138.8	13.9	1462	10	AG441877	AG441877 Mus muscu
27	c	138.6	13.8	1628	10	CG757066	CG757066 P052-2-A0
28	c	138.6	13.8	1798	10	AG171124	AG171124 Pan trogl
29	c	138.4	13.8	1885	10	AG363563	AG363563 Mus muscu
30	c	137.4	13.7	1247	10	AG035314	AG035314 Pan trogl
31	c	137.2	13.7	1355	10	AG429718	AG429718 Mus muscu
32		136.4	13.6	1281	2	BG852363	BG852363 1024034A0
33	c	136.4	13.6	1542	10	AG032943	AG032943 Pan trogl
34		136.2	13.6	1221	10	AG058113	AG058113 Pan trogl
35		136.2	13.6	1392	10	AG390642	AG390642 Mus muscu
36		136.2	13.6	2243	10	AG381986	AG381986 Mus muscu
37	c	136	13.6	1956	10	CG754548	CG754548 P050-1-A1
38	c	136	13.6	2017	10	AG435283	AG435283 Mus muscu
39		135.4	13.5	1341	10	AG030611	AG030611 Pan trogl
40		135.2	13.5	1452	10	AG032979	AG032979 Pan trogl
41	c	135	13.5	1516	2	BG809984	BG809984 mgct002xd
42	c	135	13.5	1713	10	AG435689	AG435689 Mus muscu
43	c	134.8	13.5	1341	10	AG030611	AG030611 Pan trogl
44		134.8	13.5	1577	7	CK419311	CK419311 AUF_IpOva
45		134.6	13.4	1070	9	BZ557810	BZ557810 pacs1-60_

ALIGNMENTS

RESULT 1
AG363563
LOCUS AG363563 1885 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-165G04.TJ, genomic survey sequence.
ACCESSION AG363563
VERSION AG363563.1 GI:47974768
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1885)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1885

```
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-165G04.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match           15.2%;  Score 152;  DB 10;  Length 1885;
  Best Local Similarity   49.4%;  Pred. No. 7.6e-15;
  Matches  482;  Conservative    0;  Mismatches  485;  Indels    9;  Gaps    3;

Qy         17 GCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGT 76
              |||||  |    |   | || ||||  | |  ||  |||||     | | |||   | |
Db        884 GCGGGCGGGGCGCGCGGCGGCCCACCGCGGGCGCCGGCCGCGGCCCCGGCGCCGCCCCGC 943

Qy         77 CGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGA 136
              | ||  |  || ||     | ||  |  ||   |  |   |||   |  |    || |
Db        944 CCCGCCGCCCGCGCGCGCGCCCGCCCCGGCCGCGGCCCCCGCGGCGCGGCGGCGGCGGCG 1003

Qy        137 CCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCA 196
              |  ||    | || | |||||||| || ||  ||   || ||||| |||||    ||| |
Db       1004 CGGCGGGCCGCGCGCCGGCCCGCCCGCCGCCCGCGCCGCGCGCGCCCCGGGCGCCGCCCC 1063

Qy        197 GGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGC 256
              |  ||||  | |||||| | ||    |   |  |  | ||  | | ||| ||  | ||
Db       1064 CGCCGCGCCGCGGCCCGGGCGCGCCGGCGCGGCCCCGCGCGCGGGCGGCGGCGCGCCCGG 1123

Qy        257 CCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGT 316
              |||   |||    | |||| |||        |  |      ||  || ||| |||||
Db       1124 GTGCGCCCGCCCGCGGCGCCGGGCGCCGCCGGGGGCCGCGCGGCGCCGCGGGCCGCCGGC 1183

Qy        317 C--GAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCG 374
              |  | ||||||||  || || |  ||| |   |   ||  ||    ||||| |  || | |
Db       1184 CCGGGGCGGCGCCACGCCGCGCGCGCCCGGCCGCGCCGCGCGCGGCGCCGCCCGGGGGCC 1243

Qy        375 GTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGC 434
              |  |      |||||||   | | || || ||    || || |   | ||| ||||
Db       1244 AGCCGCCCGCGGTGACGCGCCGCACGCGGCCGCGCCGGCCGCCGCGGGTCCGCGCGGCGC 1303

Qy        435 GAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCG 494
              || | | ||    |  |||  || | |          |    | ||||| || ||
Db       1304 CCGCGCGGCGC----CGTGCGCGCCACGCGCGGCCCGCGCGGGCCGGCGCCGCCGCGCGC 1359

Qy        495 CGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGA 554
              || | |||    |      |  ||||| |  |||  |||  |    |||       ||| |
Db       1360 CGCCCCGCGCGGACGCCGCGGCCGCGCGCGCCCACGGCCCGGGGCGGGCGCGCCGCGGCG 1419

Qy        555 CGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCAC 614
              |||  | ||  | | |  |   ||||| |||  |    |  |||  | | |  |||  ||
Db       1420 CGCCCGGCGCGCCCCCGCGCCCGCGCGCGCGCCGAACGACGCGCCGGGCGGCGCCGGCGG 1479

Qy        615 CGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGT 674
              || |    |  |||    |  |  | |  |  | ||| |  ||     | |  || ||
Db       1480 CGCGGGCGCGCGCGCGAGCGCGCCGAGCGGACGGCGCGCGCGCGCGCGGCGCGGCCCGCG 1539

Qy        675 GGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGC 734
              | |||| | | ||  |||     |  | ||       |   |||      || | ||  |  | |
Db       1540 GCGCCGGGCCGCGGCGCCGCCGCGGCCGCCGCGGCCGCGGCGCGCCGGGGCCGGGGGCCC 1599

Qy        735 GGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGC 794
              ||  |  |||| ||          |||||   | |  |  | || ||||  |||    | |
Db       1600 CGCGCCGACGCGCCCCCCGGCGGGCGCGGCGCGGCGCGGCGGCCCGGCGGCGCCGCGCCC 1659

Qy        795 CACGGTCGCGCGCTCCACCGAGGGGCCGATCTC---GCGGCGCGTGCACAGGCGGGCCGC 851
              | |||   |||    | |||  ||||||  | |   ||||| |      |  ||| || ||
Db       1660 CCCGGCGCCGCCGCGCGCCGGCGGGCCGCGCGCCCAGCGGCCCCGCGGCCCGCGCGCGGC 1719

Qy        852 CTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTT 911
              ||||  |  ||  || || | | ||| || ||  | ||  | || |             ||
Db       1720 GACGCGGGCGGCCAGGGACGCCCGCCGCCGCGGCCGGCCGCGCGCGCGGCCGCGCGCGGG 1779
```

```
Qy        912 GACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGG 971
              | || || ||         ||  | |  |  || | || ||  |   | |||||  | ||
Db       1780 GGCGCCGCGCCGGCGCGCCGCGGCGCCCCGCCGCCGGCCGGCGCCGGGCGCGCGGGGCGC 1839

Qy        972 CGCGCCCGCGCCCCCG 987
              | |  |||||| || |
Db       1840 CCCCGCCGCGCGCCGG 1855


RESULT 2
AG341503/c
LOCUS       AG341503               1569 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-134O11.T7, genomic survey
            sequence.
ACCESSION   AG341503
VERSION     AG341503.1  GI:47914813
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1569)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1569
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-134O11.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match           15.1%;  Score 150.8;  DB 10;  Length 1569;
  Best Local Similarity   49.1%;  Pred. No. 1.2e-14;
  Matches  475;  Conservative    0;  Mismatches  483;  Indels   10;  Gaps    4;

Qy         41 CGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGA 100
              ||  ||||  | | |  |    |  | | |  | ||||  | |||  || |||||  |
Db       1296 CGGGCGCGCNGCGNCCGCGNGGCCCCGNCGGCGCGCCGCGCGGCGCGCGCGGCGCCGGGC 1237
```

```
Qy        101 GCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCCGAGGCGGCCCGCG 160
              |||| ||| |     |       | |  | ||    | ||||| |   ||| ||   ||
Db       1236 GCGCCCGCCGNCGCGCCGNGCGCCCGGCCGGCGNCGCGNCGCCGGCNCCGGCCGCGGCCG 1177

Qy        161 CGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGA 220
              |  ||  ||    | ||  || ||| || | |||    | |  ||  || | |||  |
Db       1176 CCCCGCCCGCCCGGCGCCGGGGCCGCGGCCCCCGCCCCGGNCCGGCGGGCGCCGCCGGGC 1117

Qy        221 GCACCTCGATCTCCTCGGCGCCCGCGC---CGGCCTCGCCCGCGGAGGCCGCGCGGGCCG 277
              |  | |   | || | || | ||||    | | | ||||     |||||   |||||
Db       1116 CCGGCCCCCGCCCCGCCGCNCGCGCGCGCGCCGGCGCCCCCGCCCCCGCCGCCGGCCCG 1057

Qy        278 CGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGT 337
              |||||     | ||  | ||  |   ||  ||        |  | |||| |   |  ||
Db       1056 CGCGGCGCCGCCGCCGCCCGCGCCCCGCGCGCCCGGCCGCCCCCCGCGCGCGCCCCCGGC 997

Qy        338 CGCGCTCCTCGG-CGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCG 396
              | |||  | ||| ||||  |   |||  |    | || |  | |   ||    |  | |
Db        996 CCCGCCGCGCGGCCGCCCGCCGCCGCGCCCGCCCCCCCGGCGCCCCCGCGCCGCCGGGNG 937

Qy        397 TCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGC 456
              | |   || ||  ||   || ||  |   | ||   ||   |  ||  | | | | ||||
Db        936 GCCGCCGCCCCGCGGCGCCCCGCGGGCCCGCGCGCGGCCCCCCGCGGNCCCCGGCGCCGC 877

Qy        457 CGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGC 516
              |    |    |   | || ||| || |    |  || || | ||||   |||  |||||
Db        876 CCGGCGCCCGCCCCCGNCCCCGCGGCCCGCCGGGCCCCCGCCGCGCCCGCCGCCGGGCGC 817

Qy        517 CGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGC 576
              | |   ||  ||  |    |    ||||||     | ||||||  |      | |  |
Db        816 GCCGCGCCCCCGCGCGGGNGGNCCCCCCGCGCGCGCCCGGGGCCGCCCGCGCGCCCGCGG 757

Qy        577 GCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAG-CACCGGGCCGACCAGCGGCGACAG 635
              |||| |||  | ||| || |   || || | || | | |||||  |  | | |  |   |
Db        756 GCGCCCGCGGCGGCCGCCCCGCNCGCCCGCCCGGGCCGCCGGGGGGCGCCGGGCGGGGCG 697

Qy        636 GTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGA 695
                | | || |||      || ||||  |  || | | |  |||| |    |   |
Db        696 CCGCCCGCGCCGCCGGCGGCGCGGCGCGGGCCGCGCCGNGCCCCGCCCCCGCGCCGCCCG 637

Qy        696 GATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAG 755
              |    || |   |  || ||||   || |   ||||||  | |  |||||  |
Db        636 CCCGGGCCGCGCCGGCGCCGCCCGCGNNGCCNCCCGGGCGGCCCGCGCCGGGCCGCCCGC 577

Qy        756 -----CCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTCC 810
                  ||||||   || | |    ||||| |  ||| ||    | ||  ||||| | ||
Db        576 GCCCCCCGCGCCCCCCNGCGCGNGCGCGCGCCGCCGCGCGCGCGCCCGCGCGCGCCCGCC 517

Qy        811 ACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCC 870
              | || |||   | |||  |||   |   |||  ||| |   |||  |         ||
Db        516 CGGGCGGCGCCCGNCGCGCCCCGCCGCCCGCCGCGCCCCGGCCGCCGCGCGCGGGGGGGG 457

Qy        871 GGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAAC 930
              ||||  | |  |  |     |  || ||| |||| |   |   | |   |      |
Db        456 GGCGCCCCCCCCCCCCCCGCCCCCGCGGGCGCCGCGGCCCGGCGNGAAAAAAAACGCCCC 397

Qy        931 GTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCCGCCGGCGCGCCCGCGCCCCCGCGC 990
              | || ||| |  |||  ||    |||||||  ||||   | ||| |||||||  | | |
Db        396 CCCCGGCGCGGCGCCGGCCCGCCCCGCGCCGCCGCCGCGCCCCCGCGCGCGCCCNCCCCC 337

Qy        991 AGGAGCGC 998
              |  |||
Db        336 CCGCCCGC 329


RESULT 3
CL078342
LOCUS       CL078342               1674 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-149D22_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-149D22, genomic survey sequence.
ACCESSION   CL078342
VERSION     CL078342.1  GI:40534255
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1674)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 300
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1. .1674
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-149D22"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match            14.9%;  Score 149.6;  DB 10;  Length 1674;
  Best Local Similarity  49.7%;  Pred. No. 1.9e-14;
  Matches  485;  Conservative    0;  Mismatches  483;  Indels    8;  Gaps    4;

Qy         24 GCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTG 83
              |||  || |  |  || ||||| |  ||  ||||||   |  | ||   ||  | |   |
Db        569 GCCCCCCCCCGGCCCCCGCCCCCCCCGCCCCGCCCCCCCCCCGCCCCCCCCGCCCCCCCG 628

Qy         84 GGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCC 143
                  | || || ||  |     | | |    |  || || |   | | || |  |
Db        629 CCGCCCCCGCCCCCCGCCCCGCGCCGCCCCCCCCCCCGCCCCCCCCCCCGCGCCCCCCGC 688

Qy        144 GCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCG 203
              ||||  ||  ||| | |  |  ||        |||||  |      ||| |       |
Db        689 CCCCGCCGCCCCCCCCCCCCCCCGCCCGCGCCGCGCGCGCGCCCCGCCGCCCCCCGCCCC 748

Qy        204 CGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGA 263
              | ||| ||| ||| ||  | |  |      | | || | ||||||  || | |||||
Db        749 CCCCCCCCCCGCCCCGCCCCCGGCCCCGCGCCCCGCCCGCGCGCCCCCCCC-CCCGCCCC 807

Qy        264 GGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGG 323
              |||||  |  |||| |   ||  |  | | |   |    ||  | ||||  |     |
Db        808 CGCCGCCCCCCCCGCCCCCCCGCCCCCCCGCCCCGGCCCCCCCCCCCGCCGCCCCCCGCG 867

Qy        324 CGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAG 383
              |  ||   ||       |  || ||   || | || || |     | |       |
Db        868 CCGCCGCCCGCCCCGCCCGCCCCCGCGCCCCCCCCGCCCCCCCCCCCCCGCCCGCCCCCC 927

Qy        384 GCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGA 443
              |   |   ||  | |  ||| |  |||  |||||  |   |  | |||||  ||  |
Db        928 CCCCCCCCGCCCCCCCGCCCCCCCCCCGCGCCCCCCGCGCCCGCCCCGCGCGCGCCGCCC 987

Qy        444 GCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCC 503
              |||  | |||||  | | |   |  | ||| | || |||   |  |||  | | | ||
Db        988 CCGCCCCGCCCGCCCCCCGCGCCCCCCCCCCCCGCCCCCCGCGCCCCCGCCCCCGCCCCCC 1047

Qy        504 GCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCC 563
              ||| || | | |||| ||  ||  |      | || |       | ||||   ||| || ||
Db       1048 CCGCCCGCGCCCCCGCCCCCCCCCCGCCCCCCGCCCCCCCCCCCCCCGCCCCGCGCCCCCCC 1107
```

```
Qy        564 GGACGCGACCGGCGC---GCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCC 620
              ||| ||| ||   || |||| || || |||||  |  || ||| || || |  |
Db       1108 CCCCGCCCCCGCGGCCCGGCCCGCCGCCCCCCCCGCCCGCCGCCCCGCCCGCCCCCGCGC 1167

Qy        621 GACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCG 680
              || | ||  | |    ||   | |||| ||  | ||||  |  || |   | |||
Db       1168 CCCCGCCCGCCGCCGCGCCCGCCCCNCCCGCGCCCCCCGCGCCCGCCGCGCCCGGCGCCC 1227

Qy        681 CGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGC 740
              |  | ||  ||||||    | |   |     |   || |||| | || |   |||| ||||
Db       1228 CCCCCCGCCGCGCGCCCGGCCCCCCCGCCCCGCCGCCCGCGCGCCCCCCCCCGCGCCCGC 1287

Qy        741 CTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGT 800
              | |||  |   ||  |||| |  |||   || ||  |||  ||  || |  |||| ||
Db       1288 CCCGCCCCGCCGCCCCCGC-CCCGCCCCCCCCGCCGCCGCCCCCCGGCCCCCCGCCCCGCC 1346

Qy        801 CGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTC 860
              ||| ||| || ||    ||||      | |  |||  ||   | ||  ||||  ||| | |
Db       1347 CGCCCGCGCCCCC---CGGCCCCCNCCCCCCCGCCCGCGNCGCCGCCCCGCGCCGCCCGC 1403

Qy        861 CTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAG 920
              | ||| | || ||   ||| || ||    ||  |||    |   ||||     | |||
Db       1404 CGCCCGGCCCCGCCCCCGCCGCCGCCCCCCGCCCCCCCNCGGCGCCGCCCGCCCCCGCGCC 1463

Qy        921 CGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGC 980
              ||   |      |||   ||  | | || ||    || |    | | || |||||  || |
Db       1464 CGCGCCCCCGCCCGGCCCCGCGCGCCGGCCGCGCCGCCCGCGCCGCCCCCGCGCGGCCCC 1523

Qy        981 GCCCCCGCGCAGGAGC 996
              || | |||  | ||
Db       1524 CGCCGCCCGCGCGCGC 1539


RESULT 4
AG171124
LOCUS       AG171124               1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match           14.9%;  Score 149.4;  DB 10;  Length 1798;
  Best Local Similarity   45.8%;  Pred. No. 2e-14;
  Matches  426;  Conservative    0;  Mismatches  496;  Indels    9;  Gaps    3;

Qy          1 CAGGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCG 60
              |  |||  ||  |     |     | ||  | ||   ||  ||||    |||  |  | ||
Db        830 CNCGCGCCGCGCGNCGCCCCCGCGGCGCGCNCGCCCGCGNCGCCGCGCGCGCNCGCCGCG 889

Qy         61 ATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGA 120
                | |||  |||  ||||     || |  ||  | |  ||      | || ||| ||
Db        890 CCGCGCGCCNCGCGCCCGCGCGNNCCGCGGCGCCGCGCCCGCCNCGCCCGCNCGGCGCNC 949

Qy        121 TCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCG 180
              | || |      |   | |||||||    |    | |   |  || || |     |||
Db        950 NNCGCCGCCCCGCGCCNCNCGCCGCCGCGGCGCNCGCCGCCCGCGCGCCG-----CCGCC 1004

Qy        181 GCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCG 240
              ||||   || |||        ||| | ||| |             |  ||  | |    | |
Db       1005 GCCCNCGGCGCGCCNGCNGCNCGCCCGCGCGCCNNNNNNGCGCCGGCGCGCGCGCGNG 1064

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300
              | ||||  ||     |  | | | |   || | ||| ||  ||  |          ||
Db       1065 CGCGCGCGCGCNNNCGCGCCCGCGNCNCNCCGCGNCGCCCGCCCGCGCGCGCGNCCGCGG 1124

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGC 360
              |  |||    ||| | || ||  |  | |  ||  | |||||  |    ||||    | ||
Db       1125 GCCGCCNCNCGCCCGCCGCGCNNCCGCNCNGCGGCGGCGCGCGCGCGCGCGCGNCGGCGC 1184

Qy        361 GCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCG 420
              ||  ||  ||||  |  |    | | |   |||||| ||    | | |||||  |  | ||
Db       1185 GCGCCGGCGCGCNCGCGGCNNCGCCCGCCGGCGCCGCCGCGCCGCGCGGCGNCGCGCGCG 1244

Qy        421 AGAGGGAGCTGCGCGAGCACCGAGCGCAGCCAC---CGCCGCAGGAGATCGAGCTCCCTC 477
              | | |  |    |  ||  | ||||  || |   |||||  | |  | |  || | | |
Db       1245 GGCGCGCNCGCGCCCGGCNGNGCGCGCNCCCGCNCGCGCCGGCCGCGCCCCCGCGCGCNC 1304

Qy        478 GCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCC 537
              || ||||   || ||   | |  |||||||| |||  || || ||   || || |  |
Db       1305 GCGGCGCGCGCGCGCGCGGGCCGCGCGCGCGCGGCGCGGGGCGCCGGCGCGCGCGGCGCG 1364

Qy        538 TCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCC 597
              |    |||| ||      |  |    | |  || ||||| |||| ||||   ||
Db       1365 CGGCGCCGCGCGGCGCCGCGGGGNCGCCGCNCCGGCGCCGCGCGCGCCNCCGCGCNCGGN 1424

Qy        598 GCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGC 657
              |    |  ||  | | | |   |  |||   |  |||| ||   |||    ||
Db       1425 CCGCGGCGGNCGNCCGGCCGCCNCNCGGCNCGCGCGGCGCCGCGGCGCGCGCGCCGGCGC 1484

Qy        658 AGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCG 717
              |||              | ||||  | |  |    |    |  |  | ||  |  | |
Db       1485 CNCGCCGCNCCNCNCCGCGCCCGCNNCNCNNCCGNNCCGCGCGCCNNCGCGNCCGCGNCN 1544

Qy        718 CGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGC 777
              |||    |  ||  ||  | |||  ||  ||||       | ||   ||  | |   | ||
Db       1545 CGCNNCNCNCCGCGCGGCGGCGCGCCGGCGCCGCCGGCGCCCGNCCGCNCGGCGCNCCGC 1604

Qy        778 GCGGCCACGCTCG-ACGCCACGGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGT 836
              | ||    | ||  || |     |||||| || |||  | ||     ||  |  ||
Db       1605 CCCGCGCGNCNCGCGCGNCGNNCGCGCGCGCGCCGCCGCCGCGCGCGCGGCGNCGGCCGG 1664

Qy        837 GCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCC 896
              | |  ||| ||   |  ||| || ||   |||  |  | | | ||  ||   |  ||
Db       1665 CCGCNCGCGCGCGCGCCNCCGCGCCGCCGCCGCCCGNCGNCCCGGCNCCGCNNGGCGCCG 1724

Qy        897 GAGCGTCGCGCCGTTGACGGCGAGCGACACG 927
                | ||       || || ||| | ||
```

```
Db       1725 CCNNCCCCCGGGCGCCGCGCCGCGCGNCGCG 1755


RESULT 5
AG393083/c
LOCUS       AG393083               1878 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-211E23.TJ, genomic survey
            sequence.
ACCESSION   AG393083
VERSION     AG393083.1  GI:48018931
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1878)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1878
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-211E23.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             14.9%;  Score 149;  DB 10;  Length 1878;
  Best Local Similarity   49.5%;  Pred. No. 2.3e-14;
  Matches  484;  Conservative    0;  Mismatches  476;  Indels   18;  Gaps    5;

Qy         32 CGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCT 91
Db       1699 CGCCGGCGCNGCTGCGCGCGCGCGCCNCGCGCGCGCGCNGCCCNGCNCCGCGCCCGCGCG 1640

Qy         92 GCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGG 151
Db       1639 CCGCGCGCCGCGCGCGCCCCCGCCGCCCGCGCGCGCGCGCCGCGNCCCGCCGCNC 1580

Qy        152 CGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCC 211
Db       1579 CGNCGCGCGCGGCGCGCCGCGCGCGCGC-GCCCGCCGCGCNGCGCGCCGCGCCGCGCC 1521

Qy        212 CGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGC 271
Db       1520 GGCGCGCGCGCGCCCGCNGCCCGCGGCCGCGCGCGNCGCCGCCCGCCGCGNCGCGCGC 1461

Qy        272 GGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCCCCGA 331
Db       1460 NGGNC-CGCGCGCGCGCGCCGGCGCGCGCCGCCGCCGCGCGCNCGCGNCCCCGCGCCCG 1402

Qy        332 CGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCG 391
Db       1401 CGCCCGCGCCGCGCNCGCCGCGCGCGCNCGCCGCGNCCCGCCCGCCGCGCGCCCGCCGCG 1342

Qy        392 CGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGCAGCC 451
Db       1341 CGCCCGNNCCGCGGCGCGCGCGCGGCGCCCCGCGCCCCCCCGCG---GCGCGCCGCCGCC 1285

Qy        452 ACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCGCGA 511
Db       1284 GCCGCGCGCCGNGCCCCGCCGCGNCCGCGCCGGCGCGGGGCCGCCCGCCCGCGCCCGCGC 1225

Qy        512 GGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACGCGA 571
Db       1224 CGCGCCGCGGCCGCGCCCCGCGCGCGCCCGCCGCCGCCGCCGCGCGCGCCCGGCCGCGCG 1165

Qy        572 CCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCGGCG 631
Db       1164 GCNGNGCCCGCCGCGCCGCGCGCCCCGCGNCCGCGCCGCCCCCCCGCCGCCGCGCCCGCC 1105

Qy        632 ACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGC 691
Db       1104 CCGCCGCCCCGCCGCCNCGCGCGCCGCGCCCGCGCGCCCCCGCCGGCCGNCCGCCGCCGC 1045

Qy        692 GCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGA 751
Db       1044 GCCCGCCGGCGGNCCCGCGGCCGCCGCCCGCGCGCNGGCCCGCCGCCGCCGCGCGCCCCC 985

Qy        752 GCAGCCGCGCGAGCCTGTCCTCCA----------GCGCGGCCACGCTCGACGCCACGGT 800
Db        984 GGCGCCCCCCGCCCGCGCCCGCCGCCCCGCCGCCGCGCCCCGCGCCGCGCGCGGCCGCGCG 925

Qy        801 CGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTC 860
Db        924 CCCCCGCCCCGCCGCCGCCCCGCCGCCGCGGCGCGGCCGC--GCGGCCCGCCCGCGCCCG 867

Qy        861 CTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAG 920
Db        866 CCCCGCCGCCCCCGCGCCCCGCNCCCGCCCGGCGCGGCCGCGCCCCGCCCGCGCGCCGCC 807

Qy        921 CGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGC 980
Db        806 CGCCGCGGCCGCCGCGCCCCGCCCCCGCCCGCCCCGCGCCGCCCCCGCCCCCCCGCCCCG 747

Qy        981 GCCCCCGCGCAGGAGCGC 998
Db        746 CCGGCCGCCCCGCCCCGC 729


RESULT 6
AG039481
LOCUS       AG039481               1299 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-016J15.F, genomic survey sequence.
ACCESSION   AG039481
VERSION     AG039481.1  GI:16568206
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
```

```
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1299)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1299
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-016J15.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             14.7%;  Score 147;  DB 10;  Length 1299;
  Best Local Similarity   50.0%;  Pred. No. 5.1e-14;
  Matches  509;  Conservative    0;  Mismatches  485;  Indels   25;  Gaps    7;

Qy          3 GGCGATGCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGAT 62
              || |  |  || | ||||         |   | |   | ||    |||   || ||  |
Db        161 GGGGGGGGGGGCGGGCGGCCGCGGCGCGCGGCGCCGCCCGGGGTCGGGNGGCGGCGGGCC 220

Qy         63 CTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATC 122
                || |   ||||  | |||||     | |||   || | | |||| | | || |||  |
Db        221 GCGGCCGCNGCGGGGGGGGTGGTTTCCCCGCGGGGCGNGGGTGCGCGGGTGGGCGCGCGC 280

Qy        123 CTCCACGAGCAGGACCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGC 182
              | |     |   | || ||| || || |           |    ||  |||
Db        281 CCCGCGNCGGGNGGCCGCGCGGCGCG-GCAAAGAAAATAATATCCCGCGTCGGCGTTTTT 339

Qy        183 CCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCC 242
              || |  || |  || | |||| | | | || | || || |  ||  | ||    ||||
Db        340 NNGGNGGCCGGGGCGGCGCCGCGGCGGGCGGGGCGCGCGCGCGGCGCGCGCCCGCCCGCC 399

Qy        243 CGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGG 302
              ||||  |||    |   | |  ||  || | |||||||||   |     ||||  |
Db        400 GCCGCCCGCCGGCGCGCGCGCGCGGCCGCGGCGCGCGGGCCCCCGCGCGCGCGCGGGN 459

Qy        303 GGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGC 362
              ||       || |  |  |  ||||| |  ||  | |||||  | ||  ||    |  |  |
Db        460 GGGGCCGCGCGGCCCCCGGCGCGCCGCCCCGGCGGCGCGCCGCGCGNGGCGCGCCGGGCC 519

Qy        363 CTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAG 422
              ||  ||| |||      | ||| |   |||    ||| || | ||||||   || ||
Db        520 GGCGGCGCGGCGGCGCCGGCGGCGGCCGGCG----CGGCGCGCCCGGCGCGCGCCGCGCC 575

Qy        423 AGGGAGCTGCGCGAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGC 482
              |       | || || |||      ||| |||||   |     |  || ||     ||
Db        576 GCGCGCGCCCCCGCGCCCCGCCGCGCGCCGCCGCCCGCGCGCCCCCCGCCCCNCGCGCGN 635

Qy        483 GCCGACGAGCCGCGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGT 542
              | | |    |   | |||  ||| |   |||| || || ||| | |  || ||   |
Db        636 CCGGCCCCCGCCGCCCNCGCGCCGCCCCNCGCGCGGCGCCCGCCGCGGCCGCGCCCGCGC 695

Qy        543 GAGCGCGCCGGACGCG----CCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCG 598
              | ||||  ||  ||||    ||||||  ||||    |||||| |||| ||||| ||| ||

Db        696 GCGCGCCGCGCGCGCGNGCCCCGCCGCGCGCGCGGCGCGCGCGCGCCGCCGCCGGCG 755

Qy        599 C--------CGTCCTCGCGAGCACCGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCG 650
              |        | | ||||      ||   || |  |||    |  |    |  || ||||
Db        756 CGCCGCGCGCCGCGCCGCGCCGNNCGCCGCGNCGCGCGNGNGCGCGCGCGGNGGCGGCCG 815

Qy        651 CTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGTCGCGCGAGATGTCGAG---CTC 707
              | |  ||  | | |       |   |  |||| | |  |||||  |    || |   | |
Db        816 CGCGCGCGCCCCCCGCGCNCGGCGCGCGCGCGCCGCCGCGCGCCCGCGCNCGCGCGCCGC 875

Qy        708 GATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTCGCGGCCGAGCAGCCGCGCGAGCCT 767
              |   || |||||||   |  ||  ||| ||    |     | || || ||  ||   |||
Db        876 GGCGCGCGCGCGCGGCGCGCCGCGCGCNGCGCGCNGCGCGNCGNGCCGCGCCGCNCCGCCG 935

Qy        768 GTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCGCTC----CACCGAGGGGCCGA 823
              | |  | |    |||   |    |||  ||| |||||||      | | | | ||| |
Db        936 GCCNNCGCGGCGCGCCCGCCGNNCCGCGGCGGNCGCGCGCGGGCGGCGCGGCGCGGCGGG 995

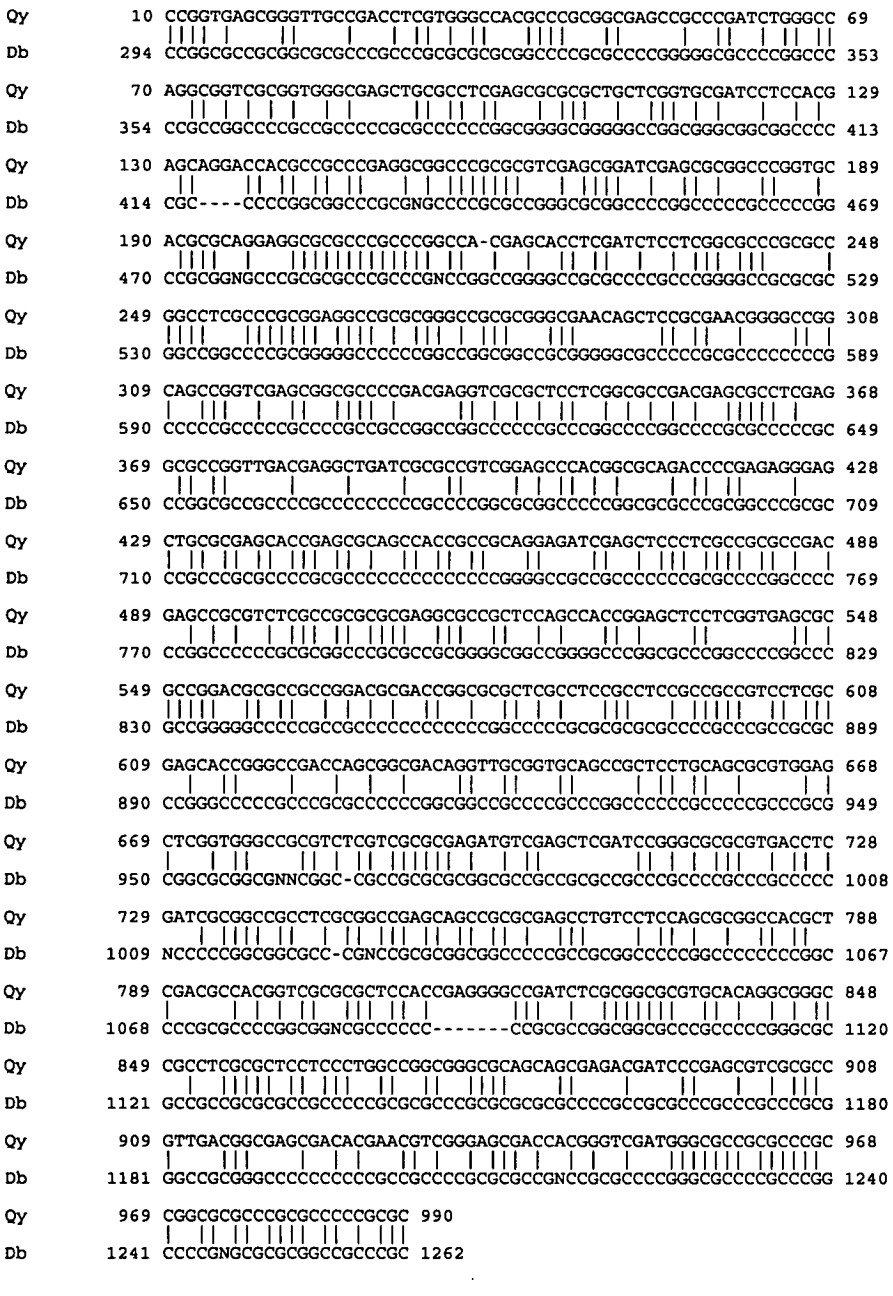
Qy        824 TCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCTGGCCGGCGGGCGCAGCA 883
                |||| |||| || | ||||   |||| |||||  || | | ||||   || ||||
Db        996 NGNCGCGCCGCGCGCCGGCGCCGCGCCGCGCGCGCCGCGCGGGCCNCCGCGCGCGCGC 1055

Qy        884 GCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACACGAACGTCGGGAGCG-AC 942
              |||  |     | || ||| ||||| |       ||  ||| | ||  |  || |  ||  |
Db       1056 GCGCGCGNCGCGCGCGCGCCGCGCGGCGCGGCGCCCGCGCCGCGCGCCGCGCGGCCGCGC 1115

Qy        943 CACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCCGCGCAGGAGCGCGAG 1001
              | |  |       | ||| | ||| || |||   |||  |   |||||||      |||| |
Db       1116 CCCNCGCGCGCGCGCGGCCCGCGCGGCGGNCGGCCGCCCGNCCCGCGCGCCCNCGCGCG 1174


RESULT 7
CG756569
LOCUS       CG756569               1406 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P051-4-B06.zb Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG756569
VERSION     CG756569.1  GI:37984265
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1406)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1406
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

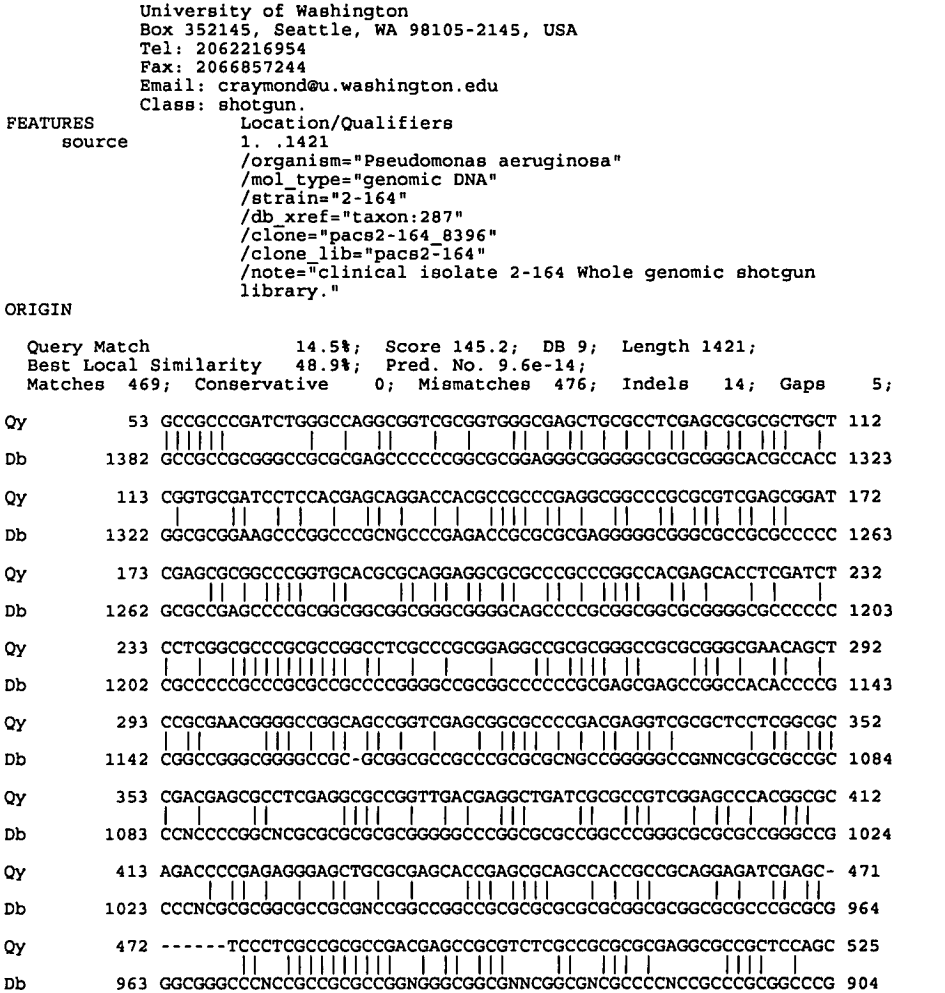
  Query Match             14.7%;  Score 146.8;  DB 10;  Length 1406;
  Best Local Similarity   50.0%;  Pred. No. 5.4e-14;
  Matches  491;  Conservative    0;  Mismatches  477;  Indels   14;  Gaps    5;
```

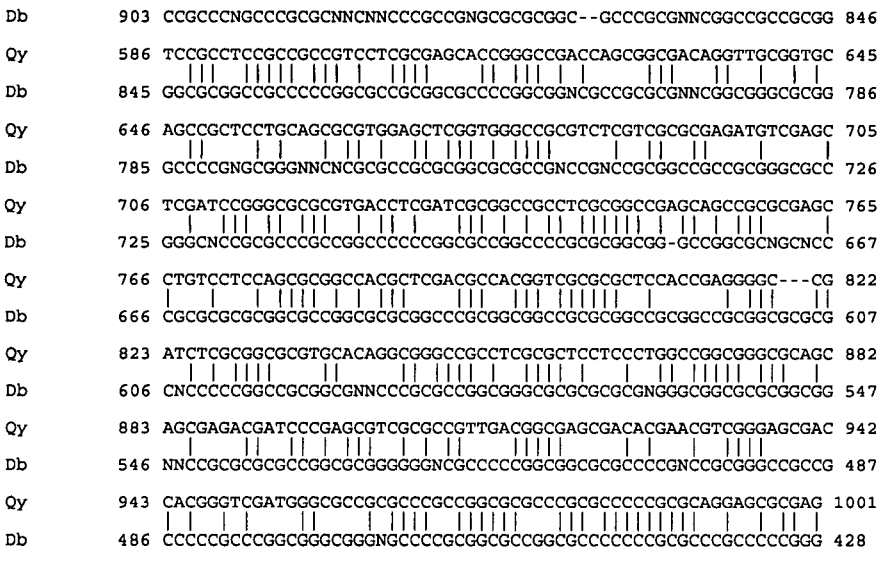


RESULT 8
BZ569488/c
LOCUS BZ569488 1421 bp DNA linear GSS 17-DEC-2002

DEFINITION pacs2-164_8396.x1 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_8396, genomic survey sequence.
ACCESSION BZ569488
VERSION BZ569488.1 GI:27204186
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1421)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .1421
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8396"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN

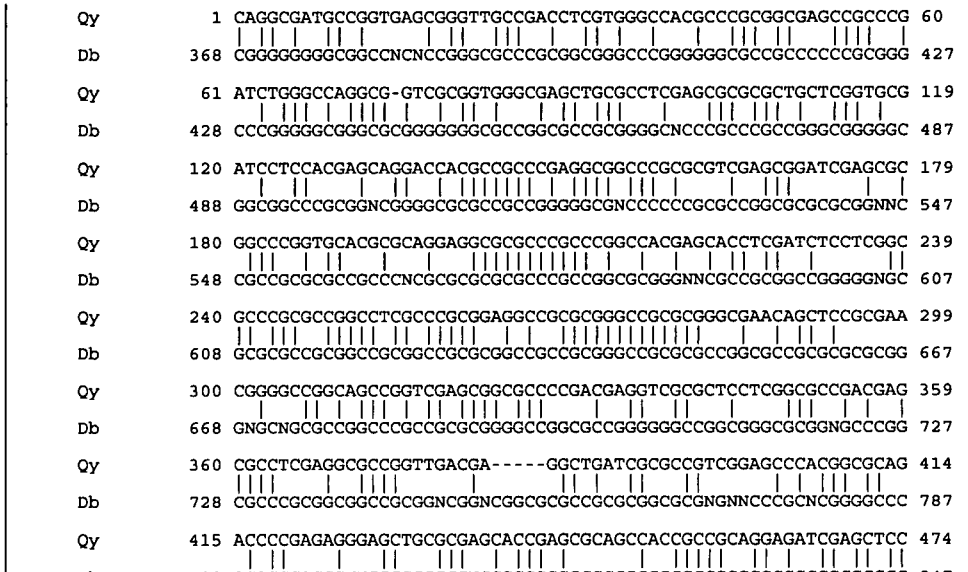
Query Match 14.5%; Score 145.2; DB 9; Length 1421;
Best Local Similarity 48.9%; Pred. No. 9.6e-14;
Matches 469; Conservative 0; Mismatches 476; Indels 14; Gaps 5;





RESULT 9
BZ569488
LOCUS BZ569488 1421 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_8396.x1 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_8396, genomic survey sequence.
ACCESSION BZ569488
VERSION BZ569488.1 GI:27204186
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1421)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .1421
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8396"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN

Query Match 14.5%; Score 145; DB 9; Length 1421;
Best Local Similarity 47.7%; Pred. No. 1e-13;
Matches 477; Conservative 0; Mismatches 510; Indels 13; Gaps 4;



RESULT 10
AG363333
LOCUS AG363333 2332 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-165A22.TJ, genomic survey

```
            sequence.
ACCESSION   AG363333
VERSION     AG363333.1  GI:47974538
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 2332)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .2332
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-165A22.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match           14.5%;  Score 144.8;  DB 10;  Length 2332;
  Best Local Similarity  50.5%;  Pred. No. 1e-13;
  Matches  486;  Conservative    0;  Mismatches  462;  Indels   14;  Gaps    5;

Qy         36 GGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCTGCGC 95
              |||| | | |||  ||  | |||||       ||  | ||  | || ||| | |     |
Db       1055 GGCCGCCCGCGCCCCGGCCGGCCCGCCGCCCCCCGCGGGGGGGGGGGGGGGGGGGCACC 1114

Qy         96 CTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGGCGGC 155
              | |    | ||| | || |  ||||  || |  || |   |   | | ||||| | |||
Db       1115 CCCCCCCCCGCCCCGCCCCCTGCGTCCCGCGGCGGGGCGCGGCGCCCCGCCGGGGGCCCG 1174

Qy        156 CCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCCCGGC 215
              || |||      | |  |   | || ||    || || |  || || | |  ||  | |
Db       1175 CCCCGCCGGCGCCCGCGCCCCCCCGCCCGCCCGCCCGAGGGGGGGGGGGGGCGGGCCCC 1234

Qy        216 CACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGC 275
              | |  || || |            |||  ||||| || |  |||| |  || |  | |||
Db       1235 CCCCCGCCCCCCCCGGGTGCGCCGGCCGCCGCCGCCCCCCGCCGCCGGCGCGGGCCCGGC 1294

Qy        276 CGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGC------AGCCGGTCGAGCGGCGCCCC 329
              ||||| ||||    |  ·|||   ||   ||| |       |||||  ||  ||||  |||
```

```
Db       1295 CGCGCCGGCGCGGGGGCGCGCCGGCGCCCCCGCCCGCGCGCGCCGGCCGCCCGGCCGCCC 1354

Qy        330 GACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGAT 389
                | ||         | | |       | || |  | |||||||    || |
Db       1355 CCCCCGGGGCGCGCGCCCCCGGCCGGCGCCGCCGGGGGCGCCGGGGCGCGGGCGGCGG 1414

Qy        390 CGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTG-CGCGAGCACCGAGCGCA 448
              | | | |  |||  |    |  ||| | ||  || | |||| || |   ||||
Db       1415 CCCCCGGCCCCCGCCGGCCCGCCGCCCGCCGCGCGGCCGCGGCCGCGGGCGCGCGGCGCG 1474

Qy        449 GCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCCGCGCG 508
              |   || | || |  | |   ||  || ||  ||| || |  || ||    | | || ||
Db       1475 GGGCCCCCGGCCGCCGGTGCGGCCGCCGCGGGGCGGCGGCCGGCGGCCCGCCCCGGCCCG 1534

Qy        509 CGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCCGGACG 568
              |   ||  || ||  ||   ||  | | |||     | || | | |  |||||| | |
Db       1535 CCGCGCCGGGCCCCGCCCGGGGGGCCCGCGCGGCCGCCCCGGCCGGCCGGCCGCCCGCGG 1594

Qy        569 CGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGACCAGCG 628
              |   ||  | |  |||  ||     ||| | || ||   || || |||   || ||  |
Db       1595 CCGGCGCGGGGGACGCGCCCCGGCGCGCGGGCGCCCGGCCGCGCGCCGCCGCGCCCCCCC 1654

Qy        629 GCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGTCTCGT 688
              ||| | |   |||  || ||||      || |||||  |   |   | ||||||    |
Db       1655 GCGCCCGCCGGCGCGGCCGCCGGCGGCGCCGCGCGCGCCCCCGCGGCGGCCGCCCGGCCC 1714

Qy        689 CGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGC--CTCGCG 746
              | |||      || |     ||| || |  |   ||  | |||   ||||  |  |||
Db       1715 CCCGCCGCCGCCCGGGGCGCGACCGCGCCCCGGGCCCCCGGGTCGGCCCCGCGGCCGGCG 1774

Qy        747 GCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCG 806
                || ||  ||| ||  ||     || ||  || | ||   | || ||| ||  |||||
Db       1775 CGCGCGCGCCCGGGCCCGCGCCCCCGCCGCCGGGCCCCTCCCCGCCGCGCGCCGCGCGCG 1834

Qy        807 CTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCT 866
              | || |||    |||   | ||||  |||  | |  ||| |||     ||| |  | | |
Db       1835 C-CCGCCGGCTCGCCCGGCGCGCGCGGCGGCCCCGCGCGCGCC----CGCCCGGCGCGCG 1889

Qy        867 GGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACAC 926
                |||   | ||| || ||| | |   ||    || | || ||  | || || | |   |
Db       1890 CCCCGCGCGCCGCCGCGGCGGGCCCGCCCGCGCCGCCCCGGCGCGGGCGCCGCGAGCGCC 1949

Qy        927 GAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCC 986
                || || | ||| |  | |         |    ||||||  |    | ||||||||  ||
Db       1950 CCGCGCCGCGCGCGCCGCCCGCCGCGCCGCGCGCGCGCCGCCGCCGGGGCCCGCGCGGCC 2009

Qy        987 GC 988
              ||
Db       2010 GC 2011
```

```
RESULT 11
CL470317/c
LOCUS       CL470317                1280 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_140_A11.v1 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_140_A11.v1, genomic survey sequence.
ACCESSION   CL470317
VERSION     CL470317.1  GI:45936028
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1280)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
```

```
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS806754; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1280
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_140_A11.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             14.3%;  Score 143.4;  DB 10;  Length 1280;
  Best Local Similarity   48.5%;  Pred. No. 1.9e-13;
  Matches  476;  Conservative    0;  Mismatches  498;  Indels    7;  Gaps    4;

Qy         17 GCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGT 76
              ||||   ||||  | ||| |  | | | |||||||||| ||||    |  |   |  | |
Db       1237 GCGGNNCGCCGCNCNCGTCGCGCGCTCGCGCGGCGAGACGCCGCGNCGCGNNCGCGAGCN 1178

Qy         77 CGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGA 136
              |   | | ||| ||   ||| ||| ||      |       |||  |   | ||  | ||
Db       1177 CCGTGCGAGCGCGCGCGGCCGCGANCGGCGCGTCGCGNCGGCGCGCGCGCGCGGCCGGGC 1118

Qy        137 CCACGCCGCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCA 196
              | || || |||  |  || ||||       ||   ||  ||||  | |     |||
Db       1117 GCGCGNCGGCCGGCGGCGCGCGCGGCCGCGCCGCCGCGTCCGCGCGCGGCGCGCCGCCGG 1058

Qy        197 GGAGGCGCGCCCGCCCGGCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGC 256
                ||    || ||  ||| || ||  | |   | |  || |  | ||| ||  | |||
Db       1057 CCGCGCCGCGCCCCCGCGCCGCGCGCGGCGCCGCCGCGGCGCCCGCGGCGGCGCGCGCGC 998

Qy        257 CCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGT 316
              ||    |  |      ||  |   | |   |  | |  ||  ||    | ||   | |
Db        997 GCGTCCCGCGCCGCANCGCGCCCGCGCCCGCCGCCGCGCCGCGCGCCCGCCGCGCGCCGN 938

Qy        317 CGAGCGGCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGT 376
              || ||||| |||       |   |||   | ||||  |  || ||| |      |   |||
Db        937 CGCGCGGCCCCCGTCGCCGCGGGCGGNACGCGGCCGCCGCGCGCGGCGGCGCGGCGCGGN 878

Qy        377 TGACG--AGGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGC 434
              | ||    || |     | |  ||| || | ||  |    | ||      | || || |
Db        877 CGCCGGCGCGCCGGCGCGGGCTCCGGCGCGCGCGNGGTNCGCGCCCGCGCCGCGCCGCCC 818

Qy        435 GAGCACCGAGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCG 494
                | |  |  | |  |   |||| | |  |  |  ||  |     |  |||  | |
Db        817 CGCCGCGGCCCCCGCCGCGCGCCTCCGCCGCGCCCGCCGCGCGCGCGCCGGCCCGCCACGGC 758

Qy        495 CGTCTCGCCGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGA 554
                ||| ||   || ||   | | |   | |||   | || | |   || |||| ||
Db        757 GCCGCCGCGGCCGCCGCGGGCGCCCGCGCCCGCCCGCGCCGCCCCCGCCCGCCCGCCCGA 698

Qy        555 CGCGCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCAC 614
              ||| ||||||  |     |||||     |||   | |     |   | |||| ||
Db        697 CGCCCCGCCGCGCCGCGGCGGCGGCGGCGCGCGCCCGCGGCCGCGGGAACGCGCGCGCGG 638

Qy        615 CGGGCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGT 674
              ||   || || ||       |   |  | || | |          ||||  |  || ||
Db        637 CGCCGCGCCCCGCCCGCCGCGCGCGGCGCGCGGGCCGCGGCCTCGCGCACGCCGCGCGCG 578

Qy        675 GGGCCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGC 734
              |   ||| | |  | ||||       | || ||  || | | |  |      | ||  |||
Db        577 CGCGNGCGCCGCCGCCCGCGCACCCGCCCGCCCGCCCCCGCCCCCCCCCCCGCGCGCCCGC 518
```

```
Qy        735 GGCCGCCTC-GCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACG 793
              |  |||| | ||| |||   | ||  |||    |     || || ||  ||   | | |    ||
Db        517 GCGCGCCCCGGCGCCCGCCCCGCGCCGCCGCCGCCNCCGCCGGCCCGCGCGCCCCGCCCG 458

Qy        794 CCACGGTCGCGCGCTCCACCGAGGGGCCGATC-TCGCGGCGCGTGCACAGGCGGGCCGCC 852
              |  | | | |   || || |||   | |||  |  | | ||||| || | | | | |||||
Db        457 CGCCCGCCCCCGGCCCCGCCGCCCGCCCGCGCGCCCCCGCGCGCGCGCCGCCCGCCCGCC 398

Qy        853 TCGCGCTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTG 912
               |||   || ||| | | | | | |   || |   | ||   |||  || | | |||  |
Db        397 CCGCCGCCCCCCCCCGCCGGCCCGNCCGCGCGGGCCGNCGCCGCCGCCCGGCCCCCCGCCG 338

Qy        913 ACGGCGAGCGACACGAACGTCGGGAGCGACCACGGGTCG---ATGGGCGCCGCGCCCGCC 969
              || ||  |  | ||  || | | |  ||    | |||| ||      | |||||| |||||
Db        337 CCGCCGCCCCGCGCGNCCGCCCGCCGCCCACGCGGGCCGCGCGCGCGCGCCGCCCCCGCG 278

Qy        970 GGCGCGCCCGCGCCCCCGCGC 990
              || ||||  | |||||||  |
Db        277 CGCCCGCCGCCCCCCCCGCCC 257
```

```
RESULT 12
AG441877/c
LOCUS       AG441877               1462 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-323H09.TJ, genomic survey
            sequence.
ACCESSION   AG441877
VERSION     AG441877.1  GI:48084940
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1462)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1462
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-323H09.TJ"
                     /sex="male"
```

/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 14.2%; Score 142.4; DB 10; Length 1462;
Best Local Similarity 49.0%; Pred. No. 2.7e-13;
Matches 474; Conservative 0; Mismatches 479; Indels 15; Gaps 4;

```
Qy         32 CGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTGGGCGAGCT 91
Db       1428 CGCGGCCGGCCGCGCNCCCCGCCCCGCCCGGCGCGCGGCGNGGCGCGCCGCGCGCCGGGCC 1369

Qy         92 GCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCCGCCCGAGG 151
Db       1368 GCGCCGCGCGCGGGCCGCGGGCGCCGCCGCCGGCGGGGCGNGGGCGGCGCCGCGGCCGGC 1309

Qy        152 CGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCGCGCCCGCC 211
Db       1308 GCGCGCGCGGGCCGCGGGCGCCGGGCGCGGCGCCCGGCCCCCGCCGCGCGGGCGGCCGNC 1249

Qy        212 CGGC----CACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGGAGGCC 267
Db       1248 CGCCGCGGCCCGGGCCGCGCCCGGCCCGGCGCGGGCGCGCCGCGGGGGGCCGGGGCCCCC 1189

Qy        268 GCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCGGCGCC 327
Db       1188 GCGCCCCGCCGGCCGGCCGGCGCGCGCGCCGCGCGCGCGCCGCGCCGGGCGGGGCC 1129

Qy        328 CCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTG 387
Db       1128 CCGGCCCNCGCCGCCCCGGCCGCGGCCCGCCGCGCGCCGCGCCGCCGGGCGCGGCCGCGG 1069

Qy        388 ATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCGAGCGC 447
Db       1068 CGCGCGCGNCGGGCCCCGGCCGCCCGCGCCCCGNCCGGGCGCGCCCCCGCCCGCGGCCG 1009

Qy        448 AGCCACCGCCGC----AGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGCC 503
Db       1008 CGCGCGCGGCGCCCCGGCGGGCGGCCGCGGCGGCGCCCCGCGCGCGCCGGGCGCCCCGCC 949

Qy        504 GCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGCCGGACGCGCCGCC 563
Db       948 CCCCCGGCCCCGGCGGGCCGGGCGCCGGGCCGCCGCCCCCGGGGCGCCCCCCGCCCGGCC 889

Qy        564 GGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGGGCCGAC 623
Db       888 GG------CGGGCCCGCCCGCCGCCCCGCCCGCCGGCCGCGCGGCCGGCGCCGCGGGCGC 835

Qy        624 CAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGGCCGCGT 683
Db       834 CGGCGGCCGCGCGGGCCCGCGCGGCCCCCGGCGCCGGCCCCCCGCCGCCCCCGCCGGCCG 775

Qy        684 CTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGCCGCCTC 743
Db       774 GCCGCCGCGCGCGGCCGCGGGCCCGGCGGCCCCCGCGGCGCGGGCGGCGCCGCGCCCCGC 715

Qy        744 GCG-GCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCG 802
Db       714 CCGCGCCGNCGCCGCGCCGCCGCGGGGCCCCGCCCGCGGCGGCGCGGGGCCCCCGGGGGC 655

Qy        803 CGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCT 862
Db       654 CGCGCCGGGGCGCGGCGCCGGGCCGGCGGGGCGCCCGCGGGGGCGCGCCCCCGCGCCCGC 595

Qy        863 CCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCG 922
Db       594 CCGGGCGNCGGGCGGGCGGCGGCGCGCCGCGCGGGCCCCCGGCCGCGCCCCCCCCCCCC 535

Qy        923 ACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGC 982
Db       534 CCCGCGCGGGGGCCCCCCNNCCCNCNCCGGGGGGGCCCCCCCCCCCNNCCCNCNCCNCCCC 475

Qy        983 CCCCGCGC 990
```

```
Db       474 CCCCCCCC 467

RESULT 13
AG448243
LOCUS       AG448243               1780 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-332C01.TJ, genomic survey
            sequence.
ACCESSION   AG448243
VERSION     AG448243.1  GI:48091306
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1780)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1780
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-332C01.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match 14.2%; Score 142; DB 10; Length 1780;
  Best Local Similarity 48.4%; Pred. No. 3e-13;
  Matches 475; Conservative 0; Mismatches 499; Indels 8; Gaps 4;

Qy         24 GCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGCCAGGCGGTCGCGGTG 83
Db       739 GCGGGCGNCGGCGCCCNCGGCCGCCGGGGGCGGGGCGGGCGGCGCCGGGNCCGCCCCCCC 798

Qy         84 GGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCACGAGCAGGACCACGCC 143
Db       799 CGCGCGC-CCGCCGCGCNCGCCCGCCCGCCNCGGGNGCCCCGCCGGGCGGGCGCGGCGCC 857

Qy        144 GCCCGAGGCGGCCCGCGCGTCGAGCGGATCGAGCGCGGCCCGGTGCACGCGCAGGAGGCG 203
```

```
Db        858 CGGNCGCCCCGCGCCGGGCGCGCCCCGCGGCCGCCGCGGCGCGGCGCCNCNGCCGCCGGCG 917

Qy        204 CGCCCGCCCG-GCCACGAGCACCTCGATCTCCTCGGCGCCCGCGCCGGCCTCGCCCGCGG 262

Db        918 CCCGCGCCCGCGCCGCCCCGCGCGCCCGCGGCCGGCCGCCCNCGCCCGCCCGCCCCGCGC 977

Qy        263 AGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAACGGGGCCGGCAGCCGGTCGAGCG 322

Db        978 GCGGGCCGCGCGGCCGNCGGCGGCGCGCCCGCCGCCCGGCCCCCGCCGGCCGCCCGGCCC 1037

Qy        323 GCGCCCCGACGAGGTCGCGCTCCTCGGCGCCGACGAGCGCCTCGAGGCGCCGGTTGACGA 382

Db       1038 GCGCGCCGGGGCGCCCCCGCGGCGCCCGCGCGGGCGGCCCCCCGCGCGCNCGGCGCGGCC 1097

Qy        383 GGCTGATCGCGCCGTCGGAGCCCACGGCGCAGACCCCGAGAGGGAGCTGCGCGAGCACCG 442

Db       1098 CCGCGCGCCGCGCGGCCCCGCCCGCNGCGCGCGCGCGCGCCCCGCGCCGCCCCCCGGCCC 1157

Qy        443 AGCGCAGCCACCGCCGCAGGAGATCGAGCTCCCTCGCCGCGCCGACGAGCCGCGTCTCGC 502

Db       1158 CCCGCGCCGGCCCCGCCGCGGGCCNGGCCNCGCGCCCCCCCGCGCCCGCCCCCCCCGCGC 1217

Qy        503 CGCGCGCGAGGCGCCGCTCCAGCCACCGGAGCTCCTCGGTGAGCGCGC-----CGGACGC 557

Db       1218 CGCGGGCGCGCCCGCGCCCCGCGCCCGCGCNGCCGCGCCCGGGCGCGCGCGGGCGCGCGC 1277

Qy        558 GCCGCCGGACGCGACCGGCGCGCTCGCCTCCGCCTCCGCCGCCGTCCTCGCGAGCACCGG 617

Db       1278 CGCCCCGGGCGCCCCCGCGGCCCNCCCCGCCGCCGGNCGGGCCGCGCCCCCCGGGCGCGG 1337

Qy        618 GCCGACCAGCGGCGACAGGTTGCGGTGCAGCCGCTCCTGCAGCGCGTGGAGCTCGGTGGG 677

Db       1338 CCGCCGCCGCCGCGCGCCGCGCGGCCCGCGGCCCCGCGCGGGCGCGCGGCCCGCGGCC 1397

Qy        678 CCGCGTCTCGTCGCGCGAGATGTCGAGCTCGATCCGGGCGCGCGTGACCTCGATCGCGGC 737

Db       1398 CCGGCCGCCCGGCCGCGGCCCGCCGCCCGGCGGGCCGCCGCCGCGCCCGCAGGCGGCGCG 1457

Qy        738 CGCCTCGCGGCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCAC 797

Db       1458 CGGCGCCGCCGCCCGCCGCGGCCCGCGGCGGCGCGCCCCCCCCGGCCGGCCGCGCGCCCC 1517

Qy        798 GGTCGCGCGCTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCG 857

Db       1518 GG-CGGGCCCGCCGCCGCGCGGGCCCCCGCCGGGCGGGCGCCGGGGCGGCCCCCCGGGCG 1576

Qy        858 CTCCTCCCTGGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGC 917

Db       1577 GCCCCCCCCGCCGCGCCGNCCCGGCCCGGNCGCGCCGCGCCCGGCCGCCCCGCCGGCGCG 1636

Qy        918 GAGCGACACGAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCC 977

Db       1637 CCGCCGCCCCCCCCACGGGCCGCGCCCCGGGGCGCCNCCGCGCCGCCCCNCCGCCCGGCCG 1696

Qy        978 CGCGCCCCCGCGCAGGAGCGCG 999

Db       1697 GGCCGCCGCGNCCCGGNCCGCG 1718


RESULT 14
AG381986/c
LOCUS       AG381986               2243 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-192D21.TJ, genomic survey
            sequence.
ACCESSION   AG381986
VERSION     AG381986.1  GI:47993191
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 2243)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .2243
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-192D21.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             14.2%;  Score 141.8;  DB 10;  Length 2243;
  Best Local Similarity   49.0%;  Pred. No. 3.1e-13;
  Matches  491;  Conservative    0;  Mismatches  490;  Indels   21;  Gaps    6;

Qy          9 GCCGGTGAGCGGGTTGCCGACCTCGTGGGCCACGCCCGCGGCGAGCCGCCCGATCTGGGC 68

Db       2097 GCGGGCGCGCGCGCGCNNCCCGGCGGCGGCGCNGNGNCNGNCGCGNCNCCNGNCCGCCGCGC 2038

Qy         69 CAGGCGGTCGCGGTGGGCGAGCTGCGCCTCGAGCGCGCGCTGCTCGGTGCGATCCTCCAC 128

Db       2037 CGCCGCGCGCGCGCCNGCNCCCCGCGCCGNGCGCGCGCCCGCCGCNCNCCNGCGCGCCGC 1978

Qy        129 GAGCAGGACCACGCCGCCCGAGGCGGCC-------CGCGCGTCGAGCGGATCGAGCGCGG 181

Db       1977 GCGCGCGCGCGCGCGGCGNGCGCCCGCCGCGCGCGCCCGCGCGCGCCCGCGCCGCGGCCG 1918

Qy        182 CCCGGTGCACGCGCAGGAGGCGCGCCCGCC-CGGCCACGAGCACCTCGATCTCCTCGGCG 240

Db       1917 CGCGCCGCCGGCGCNCNCCGCCGCCGCNGCCGCGGCCGCCGCGNGCCGGCGCCCGCCGGCG 1858

Qy        241 CCCGCGCCGGCCTCGCCCGCGGAGGCCGCGCGGGCCGCGCGGGCGAACAGCTCCGCGAAC 300

Db       1857 CCCGCCGCCGCCCCGCGCGCGGGCNCCGGCGGCCGCCGGCNCCCGCCGCGCGCCGCGCGC 1798

Qy        301 GGGGCCGGCAGCCGGTCGAGCGGCGCCCCGACGAGGT-------CGCGCTCCTCGGCGCC 353

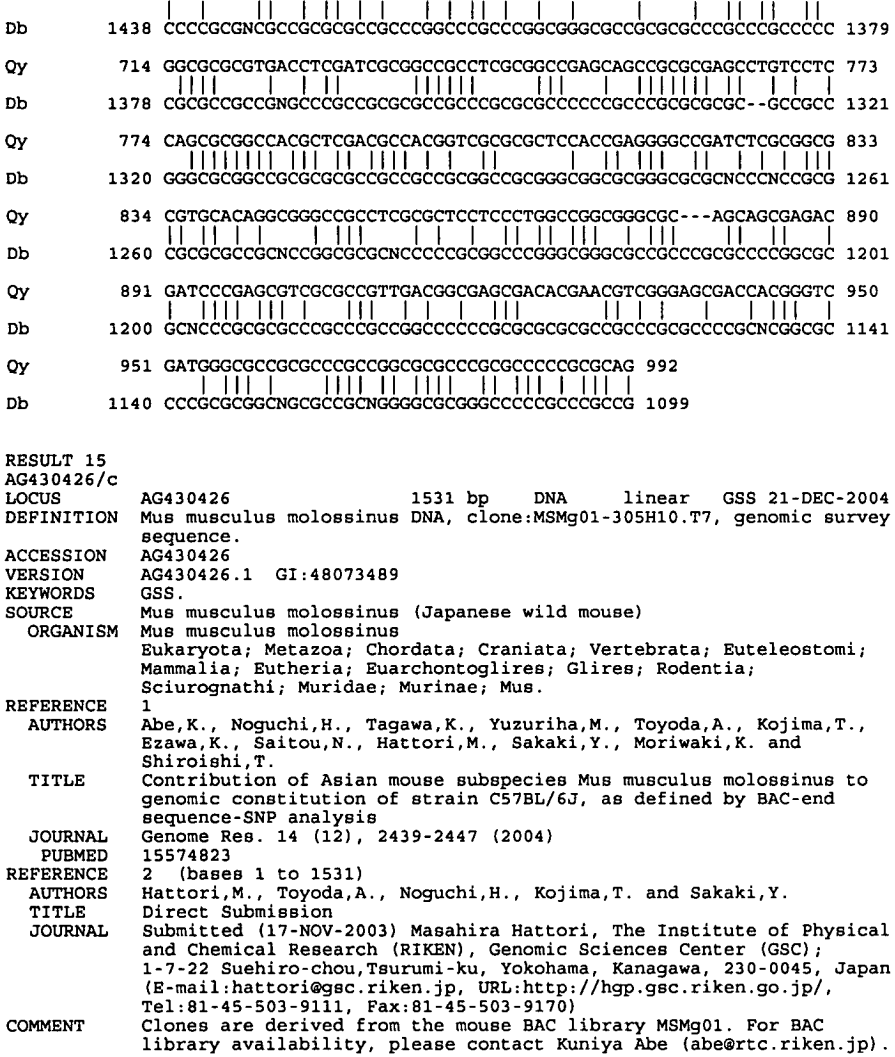
Db       1797 CGCGCCGNCGCGCCCGCCCGCGGCGCCGCGCCGCCGNCCGCCCCCCCGGCCCGCCGCCCC 1738

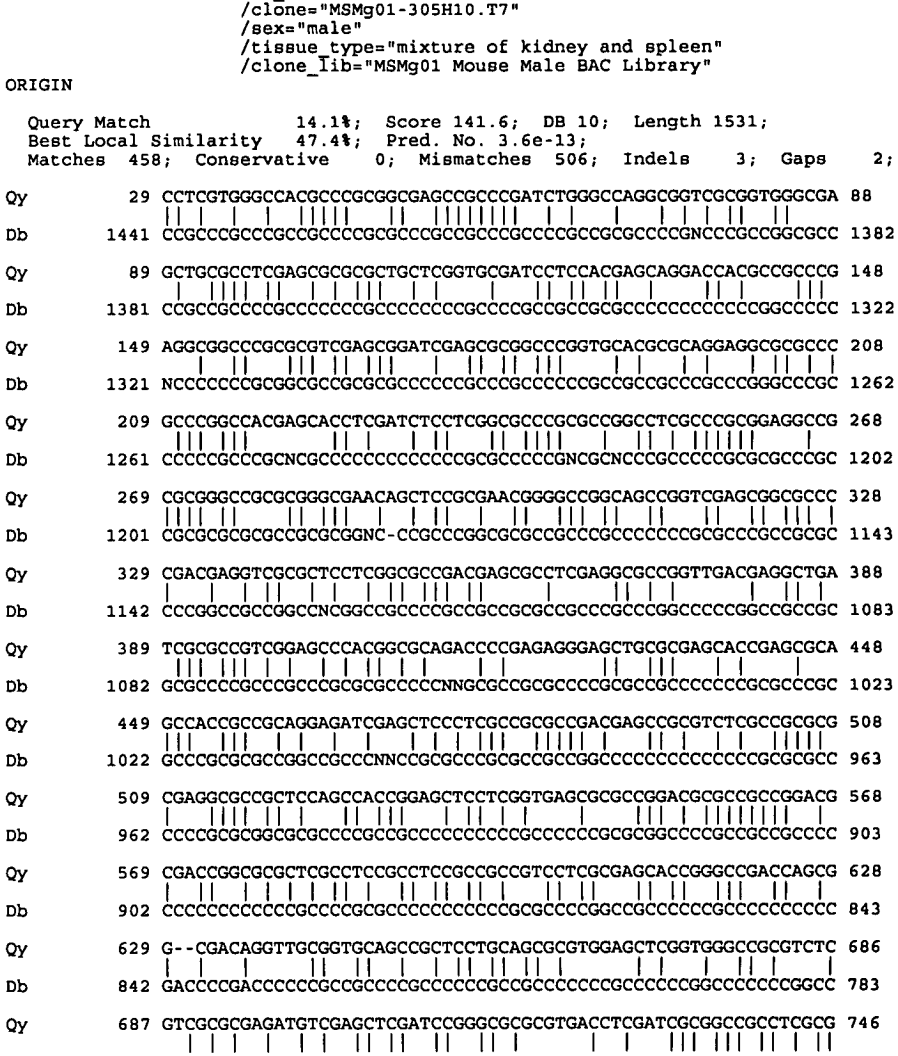
Qy        354 GACGAGCGCCTCGAGGCGCCGGTTGACGAGGCTGATCGCGCCGTCGGAGCCCACGGCGCA 413

Db       1737 GGCGCGCGCGCCCCGGCGGCGGCCGCCCGCCCCCCGCGCCNCGCGNGCGCGCCGCCGCCG 1678
```



```
RESULT 15
AG430426/c
LOCUS       AG430426               1531 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-305H10.T7, genomic survey
            sequence.
ACCESSION   AG430426
VERSION     AG430426.1  GI:48073489
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1531)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1531
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-305H10.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```





```
Db        782 CCCCCCCCCGCCGCCCCGCCCGGCCCCCGCCCCGCCCCCGCCCCGCGCCGCCCCCCCCCG 723

Qy        747 GCCGAGCAGCCGCGCGAGCCTGTCCTCCAGCGCGGCCACGCTCGACGCCACGGTCGCGCG 806
              ||  |   |||| |   ||   ||  |  ||| ||  | | || ||||   | |||||
Db        722 CCCCCGGCCCCGCCCCGCCCCCCCCCGCGCCGCCGCGCCCCGCGCCGCCCGCGGCGCGCC 663

Qy        807 CTCCACCGAGGGGCCGATCTCGCGGCGCGTGCACAGGCGGGCCGCCTCGCGCTCCTCCCT 866
              |   ||  |  ||   | | ||      ||   | |   || || | | | |  | |
Db        662 GCCGGGCGCCGCCCCCCCCCCCCCGCGNGCGGCCGCGCCCCCCCCCCGCCCCCGCGCCGCC 603

Qy        867 GGCCGGCGGGCGCAGCAGCGAGACGATCCCGAGCGTCGCGCCGTTGACGGCGAGCGACAC 926
              || ||  | | ||  || |     |||   |  ||   ||       ||   | | |
Db        602 CCGCGCCGCCCCCCGCCNCGCGCNCNCCCCCGCCCCCGGCGCGGCNGGCCCGCCNGCCGC 543

Qy        927 GAACGTCGGGAGCGACCACGGGTCGATGGGCGCCGCGCCCGCCGGCGCGCCCGCGCCCCC 986
              |  |||   |  || |  |     ||  ||| | |||||| |||  || ||| ||
Db        542 CCCCCCCGGCCCCCCCCCCCCGGGCGCGGCGGCCCCCCCCGCCCGCGCCGCCCCGCGCCG 483

Qy        987 GCGCAGG 993
              | | ||
Db        482 CCCCGGG 476
```

Search completed: April 28, 2006, 12:17:43
Job time : 4965.09 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 5541.08 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_35000_35999
Perfect score: 1000
Sequence: 1 tggactgggagccgctcgac.....acccccgcccagcggctcta 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	84868	1	DQ013294	DQ013294 Polyangiu
	2	173.6	17.4	7680	6	AR627547	AR627547 Sequence
c	3	173.6	17.4	11476	6	AR619560	AR619560 Sequence
	4	162.4	16.2	5463	6	AR627749	AR627749 Sequence
c	5	162.4	16.2	9710	6	AR619583	AR619583 Sequence
	6	137.6	13.8	25085	6	CS073978	CS073978 Sequence
	7	137.6	13.8	25085	6	AX598627	AX598627 Sequence
	8	137.6	13.8	50543	6	CS073968	CS073968 Sequence
	9	137.6	13.8	50543	6	AX598617	AX598617 Sequence
c	10	132	13.2	110000	1	CP000011_12	Continuation (13 o
	11	130.4	13.0	110000	1	BX571966_13	Continuation (14 o
	12	129	12.9	22533	6	CQ875929	CQ875929 Sequence
	13	129	12.9	161719	1	AY899214	AY899214 Streptomy
	14	129	12.9	164051	6	CQ875911	CQ875911 Sequence
	15	125.8	12.6	24081	6	CS073956	CS073956 Sequence
	16	125.8	12.6	24081	6	AX598605	AX598605 Sequence
	17	125.8	12.6	52101	6	CS073944	CS073944 Sequence
	18	125.8	12.6	52101	6	AX598593	AX598593 Sequence
	19	124.8	12.5	123580	1	AF263912	AF263912 Streptomy
	20	124.8	12.5	125401	6	AX211739	AX211739 Sequence
	21	121	12.1	110000	1	CP000076_34	Continuation (35 o
c	22	120.6	12.1	132544	1	AF521085	AF521085 Streptomy
	23	115.2	11.5	65140	6	AX211705	AX211705 Sequence
	24	114.6	11.5	23496	6	AR624244	AR624244 Sequence
c	25	114.6	11.5	26000	1	AY282522	AY282522 Myxococcu
c	26	114.6	11.5	51354	6	AR619875	AR619875 Sequence
	27	113.8	11.4	68451	1	AY330485	AY330485 Saccharop
c	28	110.8	11.1	138203	1	AY310323	AY310323 Streptomy
	29	110.6	11.1	113193	1	AF357202	AF357202 Streptomy
	30	110.6	11.1	113193	6	AX703543	AX703543 Sequence
	31	110	11.0	110000	1	CP000076_33	Continuation (34 o
	32	107	10.7	74787	1	SPA580915	AJ580915 Streptomy
	33	107	10.7	74787	6	CQ839963	CQ839963 Sequence
	34	106.4	10.6	29132	1	STH575648	AJ575648 Streptomy
	35	106	10.6	63876	1	AY834753	AY834753 Cystobact
	36	105.8	10.6	41097	1	AF016585	AF016585 Streptomy
	37	105.2	10.5	15624	6	CQ875923	CQ875923 Sequence
	38	105	10.5	16299	6	CQ875925	CQ875925 Sequence
	39	103.8	10.4	64492	1	AB086653	AB086653 Streptomy
	40	103	10.3	4893	6	AR627011	AR627011 Sequence
	41	103	10.3	15482	6	AR619672	AR619672 Sequence
	42	102.2	10.2	77534	1	AF235504	AF235504 Streptomy
	43	102.2	10.2	77536	6	BD235937	BD235937 Polyketid
	44	102.2	10.2	77536	6	AR271638	AR271638 Sequence
	45	102.2	10.2	77536	6	AR564393	AR564393 Sequence

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRAREAEREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPFFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

```
                GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
     CDS        complement(1992. .3440)
                /note="ORFu2; similar to Pseudomonas putida KT2440 PP4696;
                putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32957.1"
                /db_xref="GI:63099934"
                /translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
                APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
                QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
                KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
                EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
                VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
                ELTETEIARRLGISRKALWERRQRLGIPRPRS"
     CDS        3587. .5548
                /note="ORFu3; ORFu3p; similar to Leptospira interrogans
                AAN50474"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32958.1"
                /db_xref="GI:63099935"
                /translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
                VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
                LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
                YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
                SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
                IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
                VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
                AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
                GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
                AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
                SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
                VVAGFGATGPLRVMETLPEQYLFLTRVRP"
     CDS        complement(5702. .7231)
                /note="ORFu4; ORFu4p; similar to Burkholderia fungorum
                hypothetical protein in ZP_00033665"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32959.1"
                /db_xref="GI:63099936"
                /translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
                AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
                PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
                VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
                MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
                QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPPSANQRPASTLAQWRVGLEHP
                GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
                GRVITVHFHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
                ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
                P"
     CDS        complement(7561. .8931)
                /note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440
                PP4696; putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32960.1"
                /db_xref="GI:63099937"
                /translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
                APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
                AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
                VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
                VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
                PPPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
                IPRPRS"
     CDS        complement(8928. .11939)
                /note="ORF1; ORF1p; similar to Pseudomonas putida KT2440
                HisK (PP4695) in GenBank Accession Number AAN70268"
                /codon_start=1
                /transl_table=11
                /product="putative sensory box histidine kinase"
                /protein_id="AAY32961.1"
                /db_xref="GI:63099938"
                /translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
                YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
                AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
                FTAFFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
                HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
                FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
                VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
                LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
                IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPAREEREAAR
                LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
                LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
                RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
                LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
                GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
                DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
                DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
                IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
                LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
     CDS        complement(12405. .13379)
                /note="ORF2; ORF2p; similar to family of known or putative
                phosphotransferases, including macrolide
                2-phosphotransferases; similar to Bacillus subtilis YcbJ,
                Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273
                MphB and Staphylococcus aureus plasmid pSR1 MphBM"
                /codon_start=1
                /transl_table=11
                /product="putative macrolide kinase"
                /protein_id="AAY32962.1"
                /db_xref="GI:63099939"
                /translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
                DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
                VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
                IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
                DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
                AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
     CDS        13368. .14663
                /note="ORF3; ORF3p"
                /codon_start=1
                /transl_table=11
                /product="putative Ser/Thr protein kinase"
                /protein_id="AAY32963.1"
                /db_xref="GI:63099940"
                /translation="MFGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
                ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
                RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
                GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
                AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
                DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
                TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
                LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
     gene       15731. .33766
                /gene="dszA"
     CDS        15731. .33766
                /gene="dszA"
                /note="PKS (polyketide synthase) subunit modules 1-4a"
                /codon_start=1
                /transl_table=11
                /product="DszA"
                /protein_id="AAY32964.1"
                /db_xref="GI:63099941"
                /translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
                PRAHYSPDFNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
                IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT
```

```
                     LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                     LSPDGLCKTFDADANGYVPGDGVGVLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                     ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                     RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                     VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                     RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                     DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                     RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                     RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                     ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                     RAQFTFKKFLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                     LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                     PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                     PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                     EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE

  Query Match             100.0%;  Score 1000;  DB 1;  Length 84868;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-79;
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42574 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 42633

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42634 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 42693

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42694 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 42753

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42754 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 42813

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42814 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 42873

Qy        301 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42874 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCG 42933

Qy        361 CCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42934 CCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCG 42993

Qy        421 TGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42994 TGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG 43053

Qy        481 AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43054 AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTT 43113

Qy        541 GCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43114 GCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACC 43173

Qy        601 ACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43174 ACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGG 43233

Qy        661 AGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43234 AGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCC 43293

Qy        721 TGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43294 TGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCC 43353

Qy        781 CCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43354 CCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCT 43413

Qy        841 TCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43414 TCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCA 43473

Qy        901 CGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43474 CGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCG 43533

Qy        961 CGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 1000
              ||||||||||||||||||||||||||||||||||||||||
Db      43534 CGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 43573


RESULT 2
AR627547
LOCUS       AR627547               7680 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 8948 from patent US 6833447.
ACCESSION   AR627547
VERSION     AR627547.1  GI:59735760
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7680)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 8948 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .7680
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             17.4%;  Score 173.6;  DB 6;  Length 7680;
  Best Local Similarity   52.5%;  Pred. No. 4.8e-07;
  Matches  532;  Conservative    0;  Mismatches  459;  Indels   23;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | |||||| |  ||| | |   || |  || |||||       || ||||| || |
Db       2930 TCGACTGGGCGAAGCTGCATGGCGGCGCGCAGCTCCGGCGACGGAGCCTTCCCACCTATC 2989

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||||    |||  |||||| |  | | |    |     |||     | || || || ||
Db       2990 CCTTCGCGAAGGAGAAGTACTGGGTAGCGAAGCCGGCGCGCGCGGAGAACCGCGTCCGGG 3049

Qy        121 CGGGAGAGAGGCG---CTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGC 177
              | ||    ||     | |||| ||||| ||||| | |||| | |||    | |  | ||
Db       3050 CCGGCAGCCGGGCGGCCCTGCATCCGCTGGTGCATGAGAACTCGTCCGACCTGCGGGGGC 3109

Qy        178 ACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACG 237
              |    |||  |  ||  |||  ||| || | ||||| ||||   || ||  |||||  |
Db       3110 AGCGCTACAGCGCCCACCTCACGGGAGAGGAGTTCTTCCTCCGAGACCATGTCGTGGGCC 3169

Qy        238 GCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGG 297
              | ||| |  || ||||  | ||| |||||||||||||| ||||||||  ||  ||  |
Db       3170 AGGACAAGGTGCTGCCCGGCGTGGCCTTCCTGGAGATGGCGCGCGCCGCGGTGGTGAACG 3229

Qy        298 CGTCCGGGCGGCCGGTCCGCGACA---------TCCTCGACGTCGTCTGGACCGCGCCCG 348
              |   || | | ||    | || ||         | |   ||||||||||| |   ||||
Db       3230 CCAGTGGCCTGACGCCTCCCGGCAAGGTGACACTGCGGAACGTCGTCTGGTCACGGCCCA 3289

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGG---CCGCCGGCG 405
              || |||||  || ||  |||    || |  | ||     ||   | |||      | ||||
Db       3290 TCATCGTGACCGACGCCCCGGTGCAGGTGCACATCGGTCTGGAGGCGGAGCAGGACGGCA 3349

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              | ||| |||||  | | |   | | ||||    ||| | |   |  |   |  |||||||
Db       3350 GCGTCGTCTTCGAAATCTACGGAGGCACCGACGAGGCGCCGCTGCTCTTCTGTCAGGGGC 3409
```

```
Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
Db       3410 GGGCGCGCATCGGCGCTGGGACGGACGAGGAGGCCACTGCACCTTGCGTGGACCTGTCGG 3469

Qy        526 ACATCCTCTCACGTTGCTCG---AGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCC 582
Db       3470 TCCTCCAGTCCGGTTGTCAGGACCGGCGACAGAGCGGCGCGGCCGTCTACCAGACGTTCG 3529

Qy        583 AGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCG 642
Db       3530 CGTCGGCCGGGCTGACCTACGGACCCACGCTTCGTGCGCTCCAAGAGCTCTCGGTGGGTC 3589

Qy        643 AGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCT 702
Db       3590 GTGGCCAGGTGCTGGCGCGGGCCGTGCTCCCCGAGCAGGTGTTGAGCAGCCAGGGGCAGT 3649

Qy        703 TTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACG 762
Db       3650 TCGTCCTGCATCCGAGCTTGCTGGATGCCGCGTTGCAGGCGGCCCTCGGTCTGCACTCGG 3709

Qy        763 ATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCC 822
Db       3710 ACGAGGGAGGACAGCTCTCGCTTCCGTTCGCGCTCGATGCGCTGGAGGTGTGGGACGGGC 3769

Qy        823 CGCTGCCGGCGCGGCTC----TTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACG 878
Db       3770 CTGCTCCCTCGGACGTCTGGATTCACCTGCGCGATGAGACGGACGGCGCCGCGGGTAAGA 3829

Qy        879 AACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATG 938
Db       3830 ACCGGCAGAAGATTTC-CATCTCCCTCTGTGACCGCGACGGCAAGGTCTGGGTCCGCCTG 3888

Qy        939 CGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992
Db       3889 AAGGGATTCTCCACGCGGCCCGTGCCCGCGGCCGGTGTGCGCGCTGGCGCCGAG 3942
```

RESULT 3
AR619560/c
LOCUS AR619560 11476 bp DNA linear PAT 14-FEB-2005
DEFINITION Sequence 955 from patent US 6833447.
ACCESSION AR619560
VERSION AR619560.1 GI:59723966
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 11476)
AUTHORS Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
TITLE Myxococcus xanthus genome sequences and uses thereof
JOURNAL Patent: US 6833447-A 955 21-DEC-2004;
Monsanto Technology, LLC; St. Louis, MO
FEATURES Location/Qualifiers
source 1. .11476
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 17.4%; Score 173.6; DB 6; Length 11476;
Best Local Similarity 52.5%; Pred. No. 4.1e-07;
Matches 532; Conservative 0; Mismatches 459; Indels 23; Gaps 6;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
Db       6502 TCGACTGGGCGAAGCTGCATGGCGGCGCGCAGCTCCGGCGACGGAGCCTTCCCACCTATC 6443

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
Db       6442 CCTTCGCGAAGGAGAAGTACTGGGTAGCGAAGCCGGCGCGCGCGGAGAACCGCGTCCGGG 6383

Qy        121 CGGGAGAGAGGCG---CTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGC 177
Db       6382 CCGGCAGCCGGGCGGCCCTGCATCCGCTGGTGCATGAGAACTCGTCCGACCTGCGGGGGC 6323
```

```
Qy        178 ACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACG 237
Db       6322 AGCGCTACAGCGCCCACCTCACGGGAGAGGAGTTCTTCCTCCGAGACCATGTCGTGGGCC 6263

Qy        238 GCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGG 297
Db       6262 AGGACAAGGTGCTGCCCGGCGTGGCCTTCCTGGAGATGGCGCGCGCCGCGGTGGTGAACG 6203

Qy        298 CGTCCGGGCGGCCGGTCCGCGACA---------TCCTCGACGTCGTCTGGACCGCGCCCG 348
Db       6202 CCAGTGGCCTGACGCCTCCCGGCAAGGTGACACTGCGGAACGTCGTCTGGTCACGGCCCA 6143

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGG---CCGCCGGCG 405
Db       6142 TCATCGTGACCGACGCCCCGGTGCAGGTGCACATCGGTCTGGAGGCGGAGCAGGACGGCA 6083

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
Db       6082 GCGTCGTCTTCGAAATCTACGGAGGCACCGACGAGGCGCCGCTGCTCTTCTGTCAGGGGC 6023

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
Db       6022 GGGCGCGCATCGGCGCTGGGACGGACGAGGAGGCCACTGCACCTTGCGTGGACCTGTCGG 5963

Qy        526 ACATCCTCTCACGTTGCTCG---AGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCC 582
Db       5962 TCCTCCAGTCCGGTTGTCAGGACCGGCGACAGAGCGGCGCGGCCGTCTACCAGACGTTCG 5903

Qy        583 AGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCG 642
Db       5902 CGTCGGCCGGGCTGACCTACGGACCCACGCTTCGTGCGCTCCAAGAGCTCTCGGTGGGTC 5843

Qy        643 AGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCT 702
Db       5842 GTGGCCAGGTGCTGGCGCGGGCCGTGCTCCCCGAGCAGGTGTTGAGCAGCCAGGGGCAGT 5783

Qy        703 TTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACG 762
Db       5782 TCGTCCTGCATCCGAGCTTGCTGGATGCCGCGTTGCAGGCGGCCCTCGGTCTGCACTCGG 5723

Qy        763 ATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCC 822
Db       5722 ACGAGGGAGGACAGCTCTCGCTTCCGTTCGCGCTCGATGCGCTGGAGGTGTGGGACGGGC 5663

Qy        823 CGCTGCCGGCGCGGCTC----TTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACG 878
Db       5662 CTGCTCCCTCGGACGTCTGGATTCACCTGCGCGATGAGACGGACGGCGCCGCGGGTAAGA 5603

Qy        879 AACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATG 938
Db       5602 ACCGGCAGAAGATTTC-CATCTCCCTCTGTGACCGCGACGGCAAGGTCTGGGTCCGCCTG 5544

Qy        939 CGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992
Db       5543 AAGGGATTCTCCACGCGGCCCGTGCCCGCGGCCGGTGTGCGCGCTGGCGCCGAG 5490
```

RESULT 4
AR627749
LOCUS AR627749 5463 bp DNA linear PAT 14-FEB-2005
DEFINITION Sequence 9150 from patent US 6833447.
ACCESSION AR627749
VERSION AR627749.1 GI:59736168
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 5463)
AUTHORS Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
TITLE Myxococcus xanthus genome sequences and uses thereof
JOURNAL Patent: US 6833447-A 9150 21-DEC-2004;
Monsanto Technology, LLC; St. Louis, MO

```
FEATURES             Location/Qualifiers
     source          1. .5463
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            16.2%;  Score 162.4;  DB 6;  Length 5463;
  Best Local Similarity  51.7%;  Pred. No. 5.2e-06;
  Matches  432;  Conservative    0;  Mismatches  386;  Indels   18;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||  || |||  |||||    |   |||  || ||  | |  || ||   |||||
Db       2909 TCGACTGGCGGCGGCTGTACGCGACGCACACGCCTCGTCGGCTGCGCCTCCCGGGGTACC 2968

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              | ||| |     |||||||||       ||  |   |  |       |  |  |     |
Db       2969 CGTTCTCGAAGAAGCGTTACTCCCCGCCCGTGGTGGAGGCGCAGGCCGTCTCCACGGTGA 3028

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              || |    |    | ||||| ||||| ||    |||| ||||     |    | |||
Db       3029 GAGGCGGCGCGGTGCTTCACCCCCTCGTCCACCGCAACAGCTCCGTCCTGGCTGAGCAGC 3088

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              |||  ||||  |||||||||||||||| |||||||| || || || |  ||  |||| |
Db       3089 GCTACAGCAGCACGCTCCGGGGCGACGCCTTCTACCTGGCGGACCATCAGGTCAACGGGC 3148

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGC-----------GCCGCCG 288
              ||  |||| ||||| ||||| |  |||||||||| |||          | |||
Db       3149 GCAAAGTCCTGCCCGCCGCGGCCTATCTGGAGATGGCTCGCTTCGCGGCCAGTGACGCAT 3208

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
              ||| |   |   | | | | | | | | | || |  || | |||||||| |||||
Db       3209 GGGCGGACGAAGCGGTGGGCCGGATACCTGTCCTCAGGAACATGGTCTGGACCTCGCCCT 3268

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
              | || ||||  | || | ||    |       |      |  || ||  || ||
Db       3269 TGGTGGTGGATGACGCGGCGGAGGTCGGAGTCGCGCTTCATGCGCAGGATGCGGGGCGGC 3328

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468
              |    ||||  |   |||||  ||||    |||| |    |  ||  ||  | ||
Db       3329 TGACGTTCGAGATTCATTCCGCGCCCGACGCAGCGGGGGAGCGCACCGTCCATTGCCGCG 3388

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528
              |     ||  |  ||| | ||  |    |          ||| |  ||| || |
Db       3389 GTCTGGCGGAGTTGCAGGAGCGCGTGGCGCCGCGCCATGACCTCCGGCGGCTGGAGGCGG 3448

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588
              | | |  ||      |||    || | ||   ||| |||      | ||   | ||
Db       3449 ATTGCACCCTGCGC------GTCCTGAGCCGGGAGCAGTGTTACGAGACCTTCGCCCGGC 3502

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648
              | ||  || || | ||   | |    || |||||   |  | ||||    ||  | |||
Db       3503 TTGGTTTGAACTATGGGCCCGGTCACCGTGCGCTTGTCCGGGTGCGGGTGGGGCACGGCC 3562

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708
              | ||  | |  ||| |    || || ||||  |      |  ||| |    | || | |
Db       3563 AAGCGCTGGGCGAGGTCGAGCTGCCCGAGCACCTGGTGGGCGACGGCCGATCGTTCGTGC 3622

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGG 768
              | || ||||| ||| ||||| | || ||||||     | ||||  || ||| ||| || |
Db       3623 TGCACCCCGCGCTGGTCGACTCCGCGCTGCAAGCGGTGGTCGGTGTGTTGGGCGACGAAG 3682

Qy        769 CGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCG 824
              | |   || |||   || || ||||| || |    ||||  | | |   |||||
Db       3683 GCGGCACCGCGCCCGCGCTGCCGTTCGCAGTGGAAGAGGTCGTGGTCCATGGCCCG 3738


RESULT 5
AR619583/c
LOCUS       AR619583               9710 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 978 from patent US 6833447.
ACCESSION   AR619583
VERSION     AR619583.1  GI:59723989
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 9710)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 978 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .9710
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            16.2%;  Score 162.4;  DB 6;  Length 9710;
  Best Local Similarity  51.7%;  Pred. No. 4.1e-06;
  Matches  432;  Conservative    0;  Mismatches  386;  Indels   18;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||  || |||  |||||    |   |||  || ||  | |  || ||   |||||
Db       2556 TCGACTGGCGGCGGCTGTACGCGACGCACACGCCTCGTCGGCTGCGCCTCCCGGGGTACC 2497

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              | ||| |     |||||||||       ||  |   |  |       |  |  |     |
Db       2496 CGTTCTCGAAGAAGCGTTACTCCCCGCCCGTGGTGGAGGCGCAGGCCGTCTCCACGGTGA 2437

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              || |    |    | ||||| ||||| ||    |||| ||||     |    | |||
Db       2436 GAGGCGGCGCGGTGCTTCACCCCCTCGTCCACCGCAACAGCTCCGTCCTGGCTGAGCAGC 2377

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              |||  ||||  |||||||||||||||| |||||||| || || || |  ||  |||| |
Db       2376 GCTACAGCAGCACGCTCCGGGGCGACGCCTTCTACCTGGCGGACCATCAGGTCAACGGGC 2317

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGC-----------GCCGCCG 288
              ||  |||| ||||| ||||| |  |||||||||| |||          | |||
Db       2316 GCAAAGTCCTGCCCGCCGCGGCCTATCTGGAGATGGCTCGCTTCGCGGCCAGTGACGCAT 2257

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
              ||| |   |   | | | | | | | | | || |  || | |||||||| |||||
Db       2256 GGGCGGACGAAGCGGTGGGCCGGATACCTGTCCTCAGGAACATGGTCTGGACCTCGCCCT 2197

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
              | || ||||  | || | ||    |       |      |  || ||  || ||
Db       2196 TGGTGGTGGATGACGCGGCGGAGGTCGGAGTCGCGCTTCATGCGCAGGATGCGGGGCGGC 2137

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468
              |    ||||  |   |||||  ||||    |||| |    |  ||  ||  | ||
Db       2136 TGACGTTCGAGATTCATTCCGCGCCCGACGCAGCGGGGGAGCGCACCGTCCATTGCCGCG 2077

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528
              |     ||  |  ||| | ||  |    |          ||| |  ||| || |
Db       2076 GTCTGGCGGAGTTGCAGGAGCGCGTGGCGCCGCGCCATGACCTCCGGCGGCTGGAGGCGG 2017

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588
              | | |  ||      |||    || | ||   ||| |||      | ||   | ||
Db       2016 ATTGCACCCTGCGC------GTCCTGAGCCGGGAGCAGTGTTACGAGACCTTCGCCCGGC 1963

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648
              | ||  || || | ||   | |    || |||||   |  | ||||    ||  | |||
Db       1962 TTGGTTTGAACTATGGGCCCGGTCACCGTGCGCTTGTCCGGGTGCGGGTGGGGCACGGCC 1903

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708
              | ||  | |  ||| |    || || ||||  |      |  ||| |    | || | |
Db       1902 AAGCGCTGGGCGAGGTCGAGCTGCCCGAGCACCTGGTGGGCGACGGCCGATCGTTCGTGC 1843

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGG 768
              | || ||||| ||| ||||| | || ||||||     | ||||  || ||| ||| || |
Db       1842 TGCACCCCGCGCTGGTCGACTCCGCGCTGCAAGCGGTGGTCGGTGTGTTGGGCGACGAAG 1783

Qy        769 CGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCG 824
              | |   || |||   || || ||||| || |    ||||  | | |   |||||
```

```
Db      1782 GCGGCACCGCGCCCGCGCTGCCGTTCGCAGTGGAAGAGGTCGTGGTCCATGGCCCG 1727


RESULT 6
CS073978
LOCUS       CS073978               25085 bp    DNA     linear   PAT 05-MAY-2005
DEFINITION  Sequence 35 from Patent EP1524318.
ACCESSION   CS073978
VERSION     CS073978.1  GI:63090662
KEYWORDS    .
SOURCE      Streptomyces amphibiosporus
  ORGANISM  Streptomyces amphibiosporus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Farnet,C.M., Staffa,A., Zazopoulos,E. and Yang,X.
  TITLE     Genes and proteins for the biosynthesis of polyketides
  JOURNAL   Patent: EP 1524318-A 35 20-APR-2005;
            Ecopia Biosciences Inc. (CA)
FEATURES             Location/Qualifiers
     source          1. .25085
                     /organism="Streptomyces amphibiosporus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:221586"
ORIGIN

  Query Match             13.8%;  Score 137.6;  DB 6;  Length 25085;
  Best Local Similarity   50.1%;  Pred. No. 0.00041;
  Matches  514;  Conservative    0;  Mismatches  464;  Indels   48;  Gaps    5;

Qy          3 GACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCC 62
              |||||| | ||||    |     |||   | ||  ||  || |  |||| |||||||||
Db      23886 GACTGGCGGACGCTGTGGGGCGAACGGCACCCGCAACGGATCTCCATGCCGACGTACCCC 23945

Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
              |||   | |  ||| |||||| ||  || || |        ||    | | |||||
Db      23946 TTCCAACTCCGGCGCTACTGGCTCGACATGACCA------CCCCGGCGCACGGCCCGCAC 23999

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |       |  |  |||| ||||| |||||    |||||||||     |   || |||
Db      24000 GTCTCCCGCGGACTGCATCCGCTGGTGCACCGGAACACCTCCGACCTGAGCGAGCAGCGC 24059

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              ||| ||   |   ||   |||   | ||||||| ||||||| ||||  ||||| ||| |
Db      24060 TACACCTCGCACTTCACCGGCCGGGAGTTCTACATCGCCGACCACCGGGTGCAGGGCGAA 24119

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
              || || ||||||  |||||||| |||||||||| ||||||||||  |   | ||| |
Db      24120 CAGGTCGTCCCCGGCGCGGCGTTGCTGGAGATGGCACGCGCCGCCGCCGTCCTCGCGGCG 24179

Qy        303 GGGCGGCCGGTCCGCGAC------ATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTG 356
              ||  |  |||    ||||       | | | | || |||||| ||  |||  |    |
Db      24180 GGCGGTGCGGAGACCGACTGGGCGCTGCGCCAGGTGGTCTGGTCCCGGCCGCTGACGGCC 24239

Qy        357 GGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTC 416
              ||             || |  |        |  |  || ||| | |||||     | ||||
Db      24240 GGACGGCCCGTCGACGTGCACACCGCCGTGTCCGTGCGGGCGGACGGCGAACCGGCCTTC 24299

Qy        417 G---------CCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGG 467
              |         | | |  | ||| || ||| |||| |||  ||||  ||  | |  ||
Db      24300 GAGATCTACACGGAGGGCCCCGGCGGCGAACGCGTCGTGCACTCCACCGGACGGCTGCAC 24359

Qy        468 ATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGAC 527
                |  || || |   |  |||      ||  |  ||    |||||   ||  | |
Db      24360 CGCAGGACCGCCGGGAACGCCGCCGAACTCCTGGACGGACCCGAACTCCCCGGCGGCGCC 24419

Qy        528 ATCCTCT-----------------------CACGTTGCTCGAGGGTCACCGGCGGAGAC 563
                | ||                       |   |   |  |||  || ||  ||
Db      24420 GGACACTTGGACGTCGCCGCGCTGCGTGCCCAGTGCGACGGCACCGTCCTCGACGCCGAG 24479

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623
              | |||||| |||  ||       | || ||| |  ||||    ||   |||| ||  ||
Db      24480 GAGTGCTACGCCCGGTTCTCCGGCGTGGGCCTGGAGTACGGGCCCACGCTGCGAGCCGTC 24539


Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683
              |   |||| || | |||     |  ||||  | ||   | | || ||  |||
Db      24540 GAGACGCTGAGCGGCGGCACCCGGCAGGCGGTGGCCCGGCTGCGGCTGTCGGCCGCCGCG 24599

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743
              || |        ||   || ||||| || ||   |||||||||||| || || || | |
Db      24600 TCCGCGAGG---ACCGGGTTCGCCCTGCACCCGAGCCTGCTCGACGCCGCACTCCAGTCC 24656

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803
              |||   || ||  |   || |   |     | |    || || || ||| | | |
Db      24657 ACGGCGGGCCTCTTCACCGGTTCCGGTACGTCCTCGGCGGCCCTGCCGTTTGCCCTGGAC 24716

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863
              ||| |   | |||| ||| ||  |||  |   |   |  ||     ||        | ||
Db      24717 CGGCTGGAGGTGCTGCGCGCGACGCCCTCCTCGGGGTGGGCGGTGGCACGCTTCGCCGCC 24776

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923
              |     |             |   |   |    || ||    ||  |||   || ||||  |
Db      24777 GACGACCGCCCAGGCGGGGTGCGCCGCCTGGACATCGACGTGTGCGACGACGACGGCGAG 24836

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACC 983
              |||  | |   ||| || |  |||   |||||             |  |||| ||| ||  ||
Db      24837 GTGTGCGTACGGATCCGGGGCTTCCAGGTCCGTACGTACGGCGGCGACGCCGCCCCGTCC 24896

Qy        984 CCCGCC 989
              || ||
Db      24897 GCCTCC 24902


RESULT 7
AX598627
LOCUS       AX598627               25085 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 35 from Patent WO02088176.
ACCESSION   AX598627
VERSION     AX598627.1  GI:28398769
KEYWORDS    .
SOURCE      Streptomyces amphibiosporus
  ORGANISM  Streptomyces amphibiosporus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Farnet,C.M., Zazopoulos,E., Staffa,A. and Yang,X.
  TITLE     Genes and proteins for the biosynthesis of polyketides
  JOURNAL   Patent: WO 02088176-A 35 07-NOV-2002;
            Ecopia Biosciences Inc. (CA)
FEATURES             Location/Qualifiers
     source          1. .25085
                     /organism="Streptomyces amphibiosporus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:221586"
ORIGIN

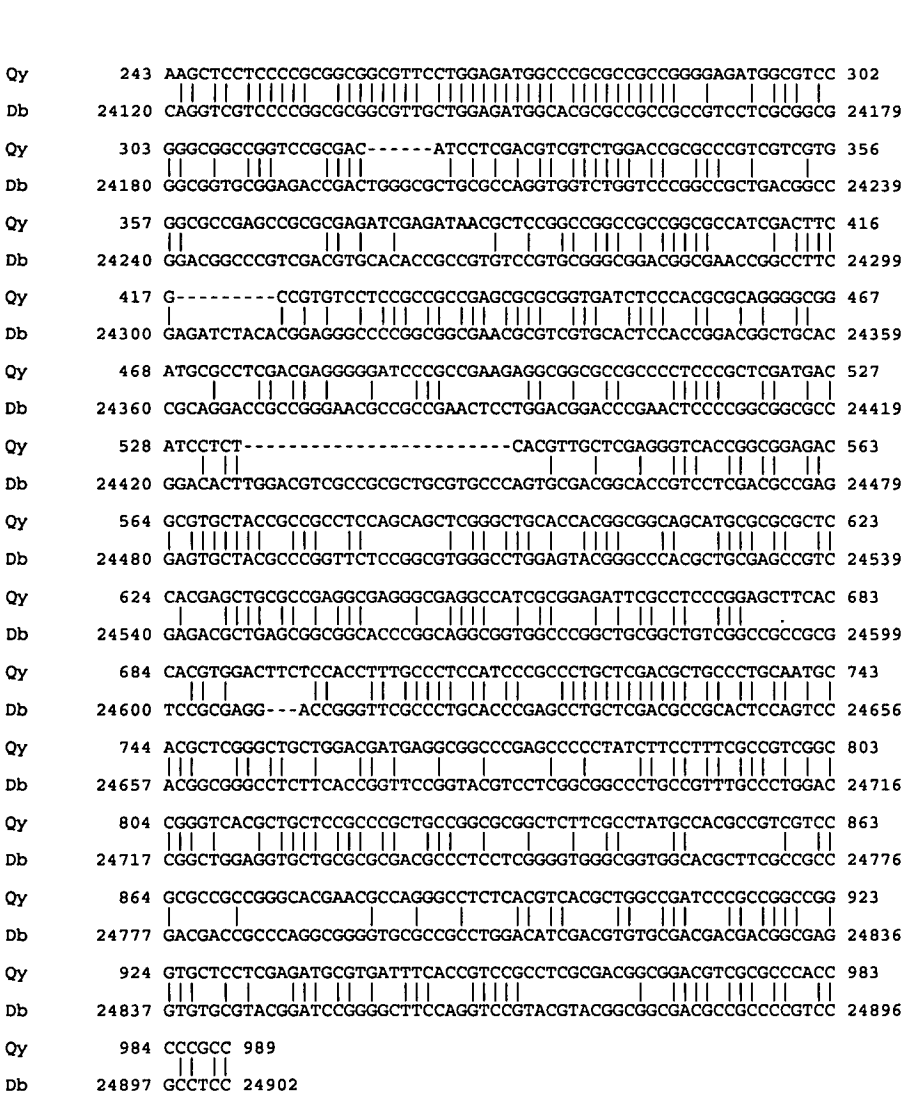
  Query Match             13.8%;  Score 137.6;  DB 6;  Length 25085;
  Best Local Similarity   50.1%;  Pred. No. 0.00041;
  Matches  514;  Conservative    0;  Mismatches  464;  Indels   48;  Gaps    5;

Qy          3 GACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCC 62
              |||||| | ||||    |     |||   | ||  ||  || |  |||| |||||||||
Db      23886 GACTGGCGGACGCTGTGGGGCGAACGGCACCCGCAACGGATCTCCATGCCGACGTACCCC 23945

Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
              |||   | |  ||| |||||| ||  || || |        ||    | | |||||
Db      23946 TTCCAACTCCGGCGCTACTGGCTCGACATGACCA------CCCCGGCGCACGGCCCGCAC 23999

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |       |  |  |||| ||||| |||||    |||||||||     |   || |||
Db      24000 GTCTCCCGCGGACTGCATCCGCTGGTGCACCGGAACACCTCCGACCTGAGCGAGCAGCGC 24059

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              ||| ||   |   ||   |||   | ||||||| ||||||| ||||  ||||| ||| |
Db      24060 TACACCTCGCACTTCACCGGCCGGGAGTTCTACATCGCCGACCACCGGGTGCAGGGCGAA 24119
```



RESULT 8
CS073968
LOCUS CS073968 50543 bp DNA linear PAT 05-MAY-2005
DEFINITION Sequence 25 from Patent EP1524318.
ACCESSION CS073968
VERSION CS073968.1 GI:63090657
KEYWORDS .
SOURCE Streptomyces amphibiosporus
ORGANISM Streptomyces amphibiosporus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Farnet,C.M., Staffa,A., Zazopoulos,E. and Yang,X.
TITLE Genes and proteins for the biosynthesis of polyketides
JOURNAL Patent: EP 1524318-A 25 20-APR-2005;
Ecopia Biosciences Inc. (CA)

FEATURES Location/Qualifiers
source 1. .50543
/organism="Streptomyces amphibiosporus"
/mol_type="unassigned DNA"
/db_xref="taxon:221586"
ORIGIN

Query Match 13.8%; Score 137.6; DB 6; Length 50543;
Best Local Similarity 50.1%; Pred. No. 0.00031;
Matches 514; Conservative 0; Mismatches 464; Indels 48; Gaps 5;

```
Db      39323 GACGACCGCCCAGGCGGGGTGCGCCGCCTGGACATCGACGTGTGCGACGACGACGGCGAG 39382

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACC 983

Db      39383 GTGTGCGTACGGATCCGGGGCTTCCAGGTCCGTACGTACGGCGGCGACGCCGCCCCGTCC 39442

Qy        984 CCCGCC 989

Db      39443 GCCTCC 39448


RESULT 9
AX598617
LOCUS       AX598617               50543 bp    DNA     linear   PAT 14-FEB-2003
DEFINITION  Sequence 25 from Patent WO02088176.
ACCESSION   AX598617
VERSION     AX598617.1  GI:28398764
KEYWORDS    .
SOURCE      Streptomyces amphibiosporus
  ORGANISM  Streptomyces amphibiosporus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Farnet,C.M., Zazopoulos,E., Staffa,A. and Yang,X.
  TITLE     Genes and proteins for the biosynthesis of polyketides
  JOURNAL   Patent: WO 02088176-A 25 07-NOV-2002;
            Ecopia Biosciences Inc. (CA)
FEATURES             Location/Qualifiers
     source          1. .50543
                     /organism="Streptomyces amphibiosporus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:221586"
ORIGIN

  Query Match             13.8%;  Score 137.6;  DB 6;  Length 50543;
  Best Local Similarity   50.1%;  Pred. No. 0.00031;
  Matches  514;  Conservative    0;  Mismatches  464;  Indels   48;  Gaps    5;

Qy          3 GACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCC 62

Db      38432 GACTGGCGGACGCTGTGGGGCGAACGGCACCCGCAACGGATCTCCATGCCGACGTACCCC 38491

Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122

Db      38492 TTCCAACTCCGGCGCTACTGGCTCGACATGACCA------CCCCGGCGCACGGCCCGCAC 38545

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182

Db      38546 GTCTCCCGCGGACTGCATCCGCTGGTGCACCGGAACACCTCCGACCTGAGCGAGCAGCGC 38605

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242

Db      38606 TACACCTCGCACTTCACCGGCCGGGAGTTCTACATCGCCGACCACCGGGTGCAGGGCGAA 38665

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302

Db      38666 CAGGTCGTCCCCGGCGCGGCGTTGCTGGAGATGGCACGCGCCGCCGCCGTCCTCGCGGCG 38725

Qy        303 GGGCGGCCGGTCCGCGAC------ATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTG 356

Db      38726 GGCGGTGCGGAGACCGACTGGGCGCTGCGCCAGGTGGTCTGGTCCCGGCCGCTGACGGCC 38785

Qy        357 GGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTC 416

Db      38786 GGACGGCCCGTCGACGTGCACACCGCCGTGTCCGTGCGGGCGGACGGCGAACCGGCCTTC 38845

Qy        417 G---------CCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGG 467

Db      38846 GAGATCTACACGGAGGGCCCCGGCGGCGAACGCGTCGTGCACTCCACCGGACGGCTGCAC 38905

Qy        468 ATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGAC 527

Db      38906 CGCAGGACCGCCGGGAACGCCGCCGAACTCCTGGACGGACCCGAACTCCCCGGCGGCGCC 38965

Qy        528 ATCCTCT----------------------CACGTTGCTCGAGGGTCACCGGCGGAGAC 563

Db      38966 GGACACTTGGACGTCGCCGCGCTGCGTGCCCAGTGCGACGGCACCGTCCTCGACGCCGAG 39025

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623

Db      39026 GAGTGCTACGCCCGGTTCTCCGGCGTGGGCCTGGAGTACGGGCCCACGCTGCGAGCCGTC 39085

Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683

Db      39086 GAGACGCTGAGCGGCGGCACCCGGCAGGCGGTGGCCCGGCTGCGGCTGTCGGCCGCCGCG 39145

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743

Db      39146 TCCGCGAGG---ACCGGGTTCGCCCTGCACCCGAGCCTGCTCGACGCCGCACTCCAGTCC 39202

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803

Db      39203 ACGGCGGGCCTCTTCACCGGTTCCGGTACGTCCTCGGCGGCCCTGCCGTTTGCCCTGGAC 39262

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863

Db      39263 CGGCTGGAGGTGCTGCGCGCGACGCCCTCCTCGGGGTGGGCGGTGGCACGCTTCGCCGCC 39322

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923

Db      39323 GACGACCGCCCAGGCGGGGTGCGCCGCCTGGACATCGACGTGTGCGACGACGACGGCGAG 39382

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACC 983

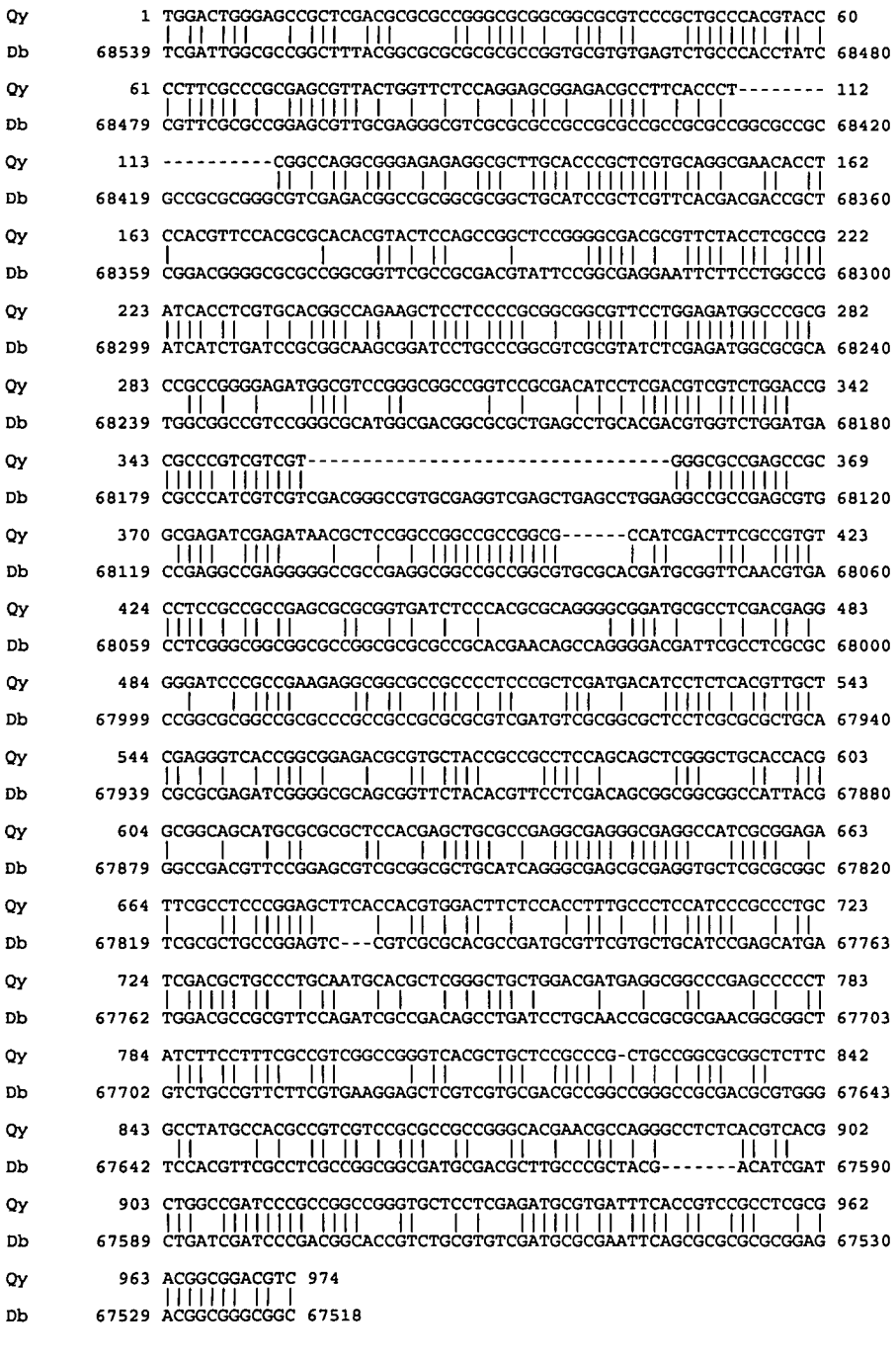
Db      39383 GTGTGCGTACGGATCCGGGGCTTCCAGGTCCGTACGTACGGCGGCGACGCCGCCCCGTCC 39442

Qy        984 CCCGCC 989

Db      39443 GCCTCC 39448

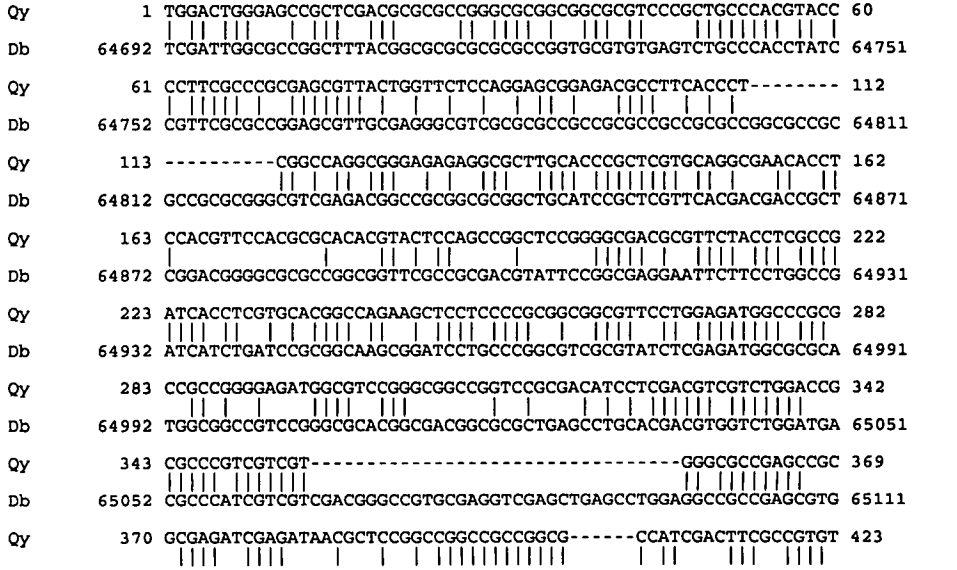

RESULT 10
CP000011_12/c
WPCOMMENT
  Sequence split into 24 fragments  LOCUS CP000011 Accession CP000011
     Fragment Name      Begin     End
     CP000011_00            1    110000
     CP000011_01       100001    210000
     CP000011_02       200001    310000
     CP000011_03       300001    410000
     CP000011_04       400001    510000
     CP000011_05       500001    610000
     CP000011_06       600001    710000
     CP000011_07       700001    810000
     CP000011_08       800001    910000
     CP000011_09       900001   1010000
     CP000011_10      1000001   1110000
     CP000011_11      1100001   1210000
     CP000011_12      1200001   1310000
     CP000011_13      1300001   1410000
     CP000011_14      1400001   1510000
     CP000011_15      1500001   1610000
     CP000011_16      1600001   1710000
     CP000011_17      1700001   1810000
     CP000011_18      1800001   1910000
     CP000011_19      1900001   2010000
     CP000011_20      2000001   2110000
     CP000011_21      2100001   2210000
     CP000011_22      2200001   2310000
     CP000011_23      2300001   2325379
Continuation (13 of 24) of CP000011 from base 1200001 (CP000011 Burkholderia mallei ATCC

  Query Match             13.2%;  Score 132;  DB 1;  Length 110000;
  Best Local Similarity   50.8%;  Pred. No. 0.0007;
  Matches  524;  Conservative    0;  Mismatches  440;  Indels   68;  Gaps    6;
```



```
RESULT 11
BX571966_13
WPCOMMENT
  Sequence split into 32 fragments  LOCUS BX571966 Accession BX571966
     Fragment Name      Begin     End
     BX571966_00            1    110000
     BX571966_01       100001    210000
     BX571966_02       200001    310000
     BX571966_03       300001    410000
     BX571966_04       400001    510000
     BX571966_05       500001    610000
     BX571966_06       600001    710000
     BX571966_07       700001    810000
     BX571966_08       800001    910000
     BX571966_09       900001   1010000
     BX571966_10      1000001   1110000
     BX571966_11      1100001   1210000
     BX571966_12      1200001   1310000
     BX571966_13      1300001   1410000
     BX571966_14      1400001   1510000
     BX571966_15      1500001   1610000
     BX571966_16      1600001   1710000
     BX571966_17      1700001   1810000
     BX571966_18      1800001   1910000
     BX571966_19      1900001   2010000
     BX571966_20      2000001   2110000
     BX571966_21      2100001   2210000
     BX571966_22      2200001   2310000
     BX571966_23      2300001   2410000
     BX571966_24      2400001   2510000
     BX571966_25      2500001   2610000
     BX571966_26      2600001   2710000
     BX571966_27      2700001   2810000
     BX571966_28      2800001   2910000
     BX571966_29      2900001   3010000
     BX571966_30      3000001   3110000
     BX571966_31      3100001   3173005
Continuation (14 of 32) of BX571966 from base 1300001 (BX571966 Burkholderia pseudomalle

  Query Match           13.0%;  Score 130.4;  DB 1;  Length 110000;
  Best Local Similarity  50.7%;  Pred. No. 0.00097;
  Matches  523;  Conservative    0;  Mismatches  441;  Indels   68;  Gaps    6;
```



```
Db      65112 CCGAGGCCGAGGGGGCCGCCGAGGCGGCCGCCGGCGTGCGCACGATGCGGTTCAACGTGA 65171

Qy        424 CCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGG 483

Db      65172 CCTCGGGCGGCGGCGCCGGCGCGCGCCGCACGAACAGCCAGGGGACGATTCGCCTCGCGC 65231

Qy        484 GGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCT 543

Db      65232 CCGGCGCGGCCGCGCCCGCCGCCGCGCGCGTCGATGTCGCGGCGCTCCTCGCGCGCTGCA 65291

Qy        544 CGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACG 603

Db      65292 CGCGCGAGATCGGGGCGCAGCGGTTCTACACGTTCCTCGACAGCGGCGGCGGCCATTACG 65351

Qy        604 GCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGA 663

Db      65352 GGCCGACGTTCCGGAGCGTCGCGGCGCTGCATCAGGGCGAGCGCGAGGTGCTCGCGCGGC 65411

Qy        664 TTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGC 723

Db      65412 TCGCGCTGCCGGAGTC---CGTCGCGCACGCCGATGCGTTCGTGCTGCATCCGAGCATGA 65468

Qy        724 TCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCT 783

Db      65469 TGGACGCCGCGTTCCAGATCGCCGACAGCCTGATCCTGCAACCGCGCGCGAACGGCGGCT 65528

Qy        784 ATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCG-CTGCCGGCGCGGCTCTTC 842

Db      65529 GTCTGCCGTTCTTCGTGAAGGAACTCGTCGTGCGACGCCGGCCGGGCCGCGACGCGTGGG 65588

Qy        843 GCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACG 902

Db      65589 TCCACGTTCGCCTCGCCGGCGGCGATGCGACGCTTGCCCGCTACG-------ACATCGAT 65641

Qy        903 CTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCG 962

Db      65642 CTGATCGACCCCGACGGCACCGTCTGCGTGTCGATGCGCGAATTCAGCGCACGCGCGGAG 65701

Qy        963 ACGGCGGACGTC 974

Db      65702 ACGGCGGGCGGC 65713


RESULT 12
CQ875929
LOCUS       CQ875929               22533 bp    DNA     linear   PAT 04-OCT-2004
DEFINITION  Sequence 36 from Patent WO2004065401.
ACCESSION   CQ875929
VERSION     CQ875929.1  GI:53789655
KEYWORDS    .
SOURCE      Streptomyces aizunensis
  ORGANISM  Streptomyces aizunensis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bachmann,B.O., Mcalpine,J.B., Zazopoulos,E. and Farnet,C.M.
  TITLE     Polyene polyketides, processes for their production and their useas
            a pharmaceutical
  JOURNAL   Patent: WO 2004065401-A 36 05-AUG-2004;
            Ecopia Biosciences Inc. (CA)
FEATURES             Location/Qualifiers
     source          1. .22533
                     /organism="Streptomyces aizunensis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:255848"
ORIGIN

  Query Match             12.9%;  Score 129;  DB 6;  Length 22533;
  Best Local Similarity   49.9%;  Pred. No. 0.0024;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60

Db       8069 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 8128

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120

Db       8129 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 8188

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180

Db       8189 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 8248

Qy        181 C-----------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225

Db       8249 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 8308

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285

Db       8309 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 8368

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345

Db       8369 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 8428

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405

Db       8429 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 8488

Qy        406 CCATCGACTTCGCCGTCGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465

Db       8489 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 8545

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525

Db       8546 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 8605

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585

Db       8606 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 8665

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645

Db       8666 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 8725

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705

Db       8726 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 8785

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765

Db       8786 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 8845

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825

Db       8846 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 8903

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885

Db       8904 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 8963

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945

Db       8964 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 9022

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985

Db       9023 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 9063


RESULT 13
AY899214
LOCUS       AY899214               161719 bp    DNA     linear   BCT 25-APR-2005
DEFINITION  Streptomyces aizunensis strain NRRL B-11277 glycosyltransferase,
            polyketide synthase type Is, acyl transferase, membrane protein,
            ABC transporter, sugar dehydratase/epimerase, sugar epimerase,
```

sugar nucleotidyltransferase, sugar dehydratase/epimerase, thioesterase, acyl CoA ligase, amine oxidase, phosphopantetheinyl transferase, transcriptional regulator, carboxylase/carboxyltransferase, amidinohydrolase, amide synthetase, 5-aminolevulinate synthase, and acyl CoA ligase genes, complete cds; and unknown gene.

ACCESSION AY899214
VERSION AY899214.1 GI:62737766
KEYWORDS .
SOURCE Streptomyces aizunensis
ORGANISM Streptomyces aizunensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 161719)
AUTHORS McAlpine,J.B., Bachmann,B.O., Piraee,M., Tremblay,S., Alarco,A.M., Zazopoulos,E. and Farnet,C.M.
TITLE Microbial Genomics as a Guide to Drug Discovery and Structural Elucidation: ECO-02301, a Novel Antifungal Agent, as an Example
JOURNAL J. Nat. Prod. 68 (4), 493-496 (2005)
PUBMED 15844935
REFERENCE 2 (bases 1 to 161719)
AUTHORS Zazopoulos,E. and Farnet,C.M.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2005) Ecopia BioSciences Inc., 7290 Frederick-Banting, Saint-Laurent, Quebec H4S 2A1, Canada
FEATURES Location/Qualifiers
source 1. .161719
/organism="Streptomyces aizunensis"
/mol_type="genomic DNA"
/strain="NRRL B-11277"
/db_xref="taxon:255848"
CDS complement(6. .1109)
/note="Orf8"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAX98183.1"
/db_xref="GI:62737767"
/translation="MTYPAIGHVFPIVPLAWALRSAGHEVLVASAGDALEAANAGLHV
ADVAPGFHLEDFLQSTAGELMARLRGPGGVDPMDGLTLFAHLNNHLADGIVRTADDFR
PDLIVFEQIFVSGLIAAARLGVPAVQHNFGFARGTQLRELTVSMLTETMARHGVDRVS
ERVPVIDIAPPSMAEPERDGWSMRPVPYNSGAVLPDWLLEKPGRRRVGVTLGTASVHI
NGLGPVQRLAAAAAGVDAEFVLALGDVDTTALGELPPNVRAVGWVPLTALLQTCDAAV
HHGGAGTTLAALNAGVPQLVLPDGADRHINAEAVRDRGAGLLGTADDLSAEVLVQLLS
DEKMTAAAREVRAEIRTMPSPVSLVPRLEELAG"
CDS 1375. .25818
/note="Orf9"
/codon_start=1
/transl_table=11
/product="polyketide synthase type I"
/protein_id="AAX98184.1"
/db_xref="GI:62737768"
/translation="MLNESEEFTPEINVASEVGGTQGESPESTPSWQQRLTGLTEAEQ
HTALLEWVSSLASAALRDAAPDTLDPHRPFLDLGFDSLAAVDLHARLVAGTGLRLPVT
LAFDHPTPAHLARHLHAAILGLTGPAETPVTAAVGSDEPIAIVGIGCHFPGGVQSPEA
LWNLVETGTDAISAFPTGRGWDLDALYDPDPDRAGTSYAREGGFLHDADAFDAAFFGI
SPREALAMDPQQRLLLEASWEAFDRAGVDPAALRGGQVGVFVGAETQEYGPRLQDATD
GFEGYLVTGNAASVASGRIAYTFGFEGPTVTVDTACSSSLAALHLAVQALRTGECSLA
LAGGVAVMASPGSFVSFSRQRGLAPDGRCKPFAAAADGTAWGEGVGMLLVERLSDARA
KGHRILAVVRGSAINQDGASNGLTAPSGPSQQRVIRQALANAGLSAAEVDVVEAHGTG
TRLGDPIEAQALLATYGQEHTDDRPLWLGSLKSNIGHTQAAAGVAGIIKMIMAMRHGV
LPRTLHVDAPTPHVDWEAGAVTLLTEAVEWPESDRPRRAGVSSFGMSGTNAHVIVEEP
AAQDREGAPTSGAQAPDSSQGQAQGTSTAPVLLPWALSAKTPEALRAQARRLGTLIAA
QPHVTPLDIGHSLATTRGRFEQRAIVLGDDREAFLDALHALAEGNDTPSVVQGAAAPG
KLAFLFTGQGSQRLGMGRELYETHPVFADALDDACWYLDDQLELPLLDVLFADEGSPE
AALLHQTAYTQPALFAVEVALFRLVDSWGLKPDFVAGHSIGEIAAAHVAGVFSLEDAC
MLVAARGRLMQALPAGGVMIALQASEDEVLPLLTDRVSIAAINGPQAVVIAGDEDAAA
AIAETFQAAGRKTKRLTVSHAFHSPHMDAMLEEFLRVAQVLDYAKPTLPVVSLLTGTT
ATPAELATPAYWVRHVRDAVRYLDGVRTLHQRGVRTFLELGPDAVLTAMAQDCVDPQG
AAFAPALRSGRPEAATVLNAVAHAHVRGAETDWAAFFAGTGAQRVDLPTYAFQRQRYW
MDSRTPAPDSAAQRAHGGADPVDRVFWDAVEHEDVATLAAALELDLDGEQPLSEVVPA
LSAWRRRRRTQSEVDGWRYRVTWKPLTEVSTSGLSGSWVVISPAGGADDSAVVSALVG
RGVDVRRVVVEAGVDRSALAGLLAEVGSPSGVVSLLGLDESGGLLGTVGLVQALGDAG
VGAPLWCLTRGAVSVGRSDRLVSPVQAQVWGLGRVAALEVPEWWGGLIDLPEVLDERA
VSRLVGVLAGSGEDQVAVRSSGVFGRRLVRAPRAEGASAWSPTGTVLVTGGTGVLGGR
VARWLAGAGAERLVLTSRRGLDAPGAVELVEELTTGFGVEVSVVACDAADRDALRALL
SAEAGSLTAVVHTAGVLDDGVLDALTPDRIDSVVRAKAVSALNLHELTAELGIELSDF
VLFSSVTGTVGAAGQANYAAANAFLDALAEQRRADGLAATSIAWGPWAEGGMAADEAM
DARMRREGMPPMAPTSAMSALEQAVGAGETALTVADIDWERFSSVIAAVRPNPLIGDF
VVGAEGTAAASGHGSVVTGADVAATVSGRLAGLTQAEQERELLSLVRLHVAAVLGHDG
SDAVGAERAFKELGFDSLTSVELRNRLGAATDLRLPTTLVYDYPTSAALAEYLRGELA
GSAQDAGPPLPAVVGSAADDDPIVIVSMSCRFPGGVRTPEDLWQLLADGTDTVAAFPA
DRGWDLDGLYSADPERSGTSYTREGGFLYDAADFDADFFGISPREALAMDPQQRLLLE
TAWETFERAGIDPASLRGSQAGVFVGTNGQDYLSLVTREGDGLDGLEGHVGTGNAASV
VSGRLSYVFGLEGPAITVDTACSSSLVALHLAVQALRQGECTLALAGGVTVMSTPDAF
VDFSRQRGLAEDGRIKAFASAADGTGWGEGVGMLLVERLSDARRNGHPVLAVVRGSAI
NQDGASNGLTAPNGPSQQRVIRQALAGAGLSAADVDAVEAHGTGTRLGDPIEAQALLA
TYGQGRPADRPLWLGSVKSNIGHTQAAAGVAGVMKMVMAMRHGVLPRTLHVDGPTPHV
DWSAGDVALLTEQREWPATGHPRRAGVSSFGLSGTNAHTIIEEAPADDDAEPTTGAGT
APSVLPLLISAKSDAGLRAQSEQLATHLVGNPDVPIGDIAYSLTTGRSGLETRAILVG
DADNRTGLAAALRSLAAGEQAPGLVQGTVTEGGLAFLFTGQGSQRLGMGRELYETYPV
FADALDAVCARMDLEVPLRDVLFGAYAGLLDETAYTQPALFAVEVALFRLVESWGLRP
DFVAGHSIGEIAAAHVAGVLSLDDACALVEARGRLMGALPGGGVMIAVQAPEAEVLPL
LTERVSIAAINGPQSVVIAGDEADAVAIVESFTGRKSKRLTVSHAFHSPHMDGMLEDF
RAVAEGLSYEAPRIPVVSNLTGALVSDEMGSAEFWVRHVREAVRFLDGMRVLEAAGVT
TYVELGPGGVLSALAQECVSGDGAAFVPVLRSGRPEAETAVTALAQAHVRGVDVDWAA
FFSGTGVQRVDLPTYAFQRQRFWPAMTAESAPVGGTVDAVDAHFWDVIEQEDVESLAE
LLGLDDASAWGSVVPALSAWRRQGQQQAQVDGWRYRASWKPVTAAVSSGVVSGTWVVA
VPAGSAGDDARVEAVTNGLAGRGVDVRRVVVEAGVDRAALAGLLAGEGSLAGVVSLLG
LDESGGLAATAGLVQALGDAGVSAPLWCLTRGAVSVGRSDRLVSPVQAQVWGLGRVAA
LEVPERWGGLVDLPEVLDERAVSRLIGVLAGSGEDQVAVRSSGVFGRRLVRAPRAEGA
ASWTPTGTVLVTGGTGVLGGRVARWLAGAGAERLVLTSRRGLDAPGTAELVEELTSSG
VEVSVVACDAADRDALRALLSSEAGSLTAVIHTAGVLDDGVLDALTPDRIDGVVRAKA
VSALNLHELTAELGIELSAFVLFSSMSGTVGTAGQANYAAANAYLDALAEQRRADGLA
ATSIAWGPWAEGGMAADAALEARMRRDGVPPMPADPAIRALRQAVAGDDAVLTVADVE
WDRFLPGFVAARHSELFSELRDVRDARAAQDRAQAAVAADRPDSLSGRLSAQAPAEQE
RELLDLVRTQVAAVLGHAGVENVGAGRAFKELGFDSLMAVELRNRIGSATELRLPATL
IYDHPTSAALAEFLRGELVGTVRVADKVLPAVVSADEDPIAIVSMSCRFPGGVRTPED
LWRLLVDGTDAVGAFPADRGWDLDRLYSPDPDQPGTSYTREGGFFDGAADFDPGFFGI
SPREALAMDPQQRLLLETSWEAIERAGIDPSSLRGSQAGVFVGTNGQDYLSLITRESE
GLEGHLGTGNAGSVMSGRVSYVLGLEGPAVTVDTACSSSLVALHWAIQALRQGECSMA
LAGGVTVMSTPENFVDFSRQRGLAEDGRIKAFASAADGTGWGEGVGMLLVERLSDARR
NGHPVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALASAGLSAADVDVVEAHGTG
TKLGDPIEAQALLATYGQDRPAGRPLWLGSIKSNIGHTQAAAGVAGIIKMVLAMQHGV
LPQTLHVDEPTPHVDWSAGEVTLLTEQTAWPTVDRPRRAGVSSFGISGTNAHTIIEQA
PAVEQLADGDATPATPALALPLPYVLSAKSPEALRAQASVLRTHLEATDHNGPGSDDL
AFSLATARAHLEHRAVLTADDPQEFREALARLADGDPSPRITTGAVSDGRTAFLFTGQ
GSQRLGMGRELYEAYPVFADALDAVCAHVDAHLEVPLKDVLFGADAGLLDQTAYTQPA
LFAVEVALFRLVESWGVKPDFVAGHSIGEIAAAHVAGVFSLQDASELVFARGRLMQAL
PTGGVMIAVQASEDEVLPLLTDRVSIAAINGPQSVVIAGDEADAVAIAESFTDRKSKR
LTVSHAFHSPHMDGMLDAFREIAEGLSYEPSRIPVVSNLTGALVSDEMGSAEFWVRHV
REAVRFLDGIRTLEAAGVTKYVELGPDGVLSAMAQDCVSGEGSVFIPVLRKARPEAES
VTTALASAHVHGIPVDWQAYFAGTGAQRVDLPTYAFQRQRYWPSAAAFVTGDPTAIGL
GDAGHPLLGAAVALADSEGVLFTGRLSLDTHPWLADHTILGSVLLPGTAFVDLAIRAG
DQVGCDVVEELTLEAPLVVPQRGGVQLQLVVEAPSGPGQRPFSVHSRRQDAYAEEPWM
RHASGVLTSGVSRRELSVEGGEFEALAVWPPTGAVPVDVRGLYEELAEAGVAYGPLFQ
GLKAAWRRDGELFTEVALPGEARREAARFGLHPALLDAGLHAIGHGEGPEPAMTGALL
PFSWAGVSLYAAGASSLRMRLTPHTPDDAHTMALLVADETGRPVAAVESLILRTASAD
QVRAADGGHLDSLFKVEWLPVAGGATPHGDSTGRRWAVLGRDGLGLPATGVQGQVAEY
DDASALGAALAAGEPVPDAVFVHPGALPGQDTDTTAASVHAAVTDALSFVQEWLADER
FAATRLVWLTSGAVADEPGAGVRDLAGSAVRGLLRSAQSENPGQLLMLDLDQDPASLA
ALPAALAAGEPELAIRRGELRTPRLTRVPSADAAAEPLGTLGDPSGTVLVTGATGTLG
GLFARHLVTAYGVRRLLLTSRRGPEAEGAAELVAELEQLGAHVELVACDAADRSALAA
LLGAVPSEHPLTAVVHTAGVLDDGILSSLTPERVAAVLRPKVDAAWNLHELTRELGLS
AFVLFSGAAAAFGAAGQGNYAAANSFLEALAEQRRAEGLPATSLAWGLWAPQTGGMAQ
QLDEVDLRRIARDGVGGLSGDEGLGLFDTAMTVDAAVLLPMRLDLAVARAQAVSTGET
PALLRALIRVPARRAVEQRTAADGASPLAARLSALPDAERERDMLLDLVCGRVAEVLGH
TDARAVDADRAFKELGFDSLTAVELRNVLKAATGLRLSPTLVFDYPTPVALARHLLAE
LAGTADDQDAVRGRKAPARPATAAVTSVTGEDPIVIVGMGCRFPGGVRSPEDLWQLVA
TGGDGITGFPSDRGWNVEALYHPDPDHAGTSYTREGGFLHDAADFDPGFFGISPREAL
AMDPQQRLLLETSWEAFERAGIDPATLRGSRTGVFAGVMYHDYVTGIGDGGSAVELPE
GVEGYLGTGNAGSIASGRIAYTFGLEGPAVTVDTACSSSLVALHWAIQALRSGECTMA
LAGGVAVMATPETFVDFSRQRGLSADGRCKSFAAAADGTGWAEGAGMLLVERLSDAER
NGHPVLAVVRGSAINQDGASNGLTAPNGPSQQRVIREALASADLSAADIDAVEAHGTG
TRLGDPIEAQALLATYGREREAGRPLWLGSIKSNIGHTQAAAGVAGIIKMVMAMRHGV
LPQTLHVDEPSPQVDWEAGEVSLLTGAMPWPQTGRPRRAGVSSFGISGTNAHTIIEQP

```
                     PTREVTPTVPVAPVVPTVPTVPVVPWVLSGKGEEALRAQARQLQSYVLRAPELRPVDI
                     AGSLAVGRASFEDRAAVVAADREGLLAALAALADGGSATGAVEGSAVGGKLAFLFTGQ
                     GSQRLGMGRELYEAYPVFAEALDAVCARLELPLKDVLFGADAGLLDETAYTQPALFAV
                     EVALFRLVESWGLRPDFVAGHSIGEIAAAHVAGVFSLDDACALVEARGRLMGALPAGG
                     VMIAVQASEDEVLPLLTDRVSIAAINGPRSVVIAGDEADAVAIVESFTGRKSKRLTVS
                     HAFHSPHMDGMLEDFRAVAEGLSYEAPRIPVVSNLTGTLVTDEMGSAEFWVRHVREAV
                     RFLDGIRALEAAGVTTYVELGPGGVLSALAQECVSGDGAAFVPVLRSGRSEAETAVTA
                     LAQAHVRGVNVDWAAFFAGTGAERVDLPTYAFQRQRYWLHIPRVAQSGVADEVDARFW
                     DAVEREDLESLASTLEVDDESAWSSVLPALSAWRRERRAQSEVDGWRYRVSWKPLAEV
                     SASGLSGSWVVISPAGSVDDSAVVSALVGRGAEVRRVVVEAGVDRSALAGLLADAGSA
                     AGVVSLLGLDESEGLLGTVGLVQALGDAGVEAPLWCLTRGAVSVGRSDRLVSPVQAQV
                     WGLGRVAALEVPERWGGLVDLPEVLDERAVARLVGVLAGSGEDQVAVRSSGVFGRRLV
                     RAPRAEGASAWTPTGTVLVTGGTGVLGGRVARWLAGAGAERLVLTSRRGPDAPGAAEL
                     VEELTTGPGVEVSVVACDAADRDALRTLLSAEAGTLTAVIHTAGVLDDGVLDALTPDR
                     IDSVLRAKAVSAFNLHELTAELGIELSAFVLFSSMSGTVGAAGQANYAAANAYLDALA
                     EQRRADGLAATSLAWGPWAEGGMAGDDAMDARMRREGLPPMAPDAALTLLRQSVGSAD
                     AALMVVDVEWQRFAPALTVVRPSNLLAELPEARPAGTDSRTGGATSSEGAGSFAERLA
                     ALGGAEQDKELLNLVRTHIAAVLGHGGSEAVGAERAFKELGFDSLTAVELRNRLGAAT
                     GVRLPATLIFDYPTATALAAYLRGELLGTQVVVSGPVSNGVVVDDDPIAIVAMSCRFP
                     GGVRTPEDLWRLLSTGGDAIGEFPADRGWDLSRLYSPDPDKQGTFYARAGGFLYDAAD
                     FDADFFGISPREALAMDPQQRLLLETSWEAFERAGIDPSSLRGSQAGVFVGTNGQDYG
                     AMLQTIPDGIEGFLGTGNAASVVSGRLSYAFGLEGPAVTVDTACSASLVALHWAVQAL
                     RSGECSLALAGGVTVMSSPGAYIDFSRQRGLAEDGRIKAFAAAADGTGWGEGVGMLLV
                     ERLSDARRNGHPVLALVRGSAINQDGASNGLTAPNGPSQQRVIRQALANAGLSAAEVD
                     AVEAHGTGTRLGDPIEVQALLATYGREREADQPLWLGSIKSNIGHTQAAAGVAGVIKM
                     VLAMEHGVLPQTLHVDEPTPHVDWSAGDVALLTDAVEWPETGRPRRAGVSSFGFSGTN
                     AHTVLEQAPKPEEPEESQQPEETNAPARPHQSGVMPWTLSAKSEAALRVQAERLRTRI
                     ASDPLLQPVDVAYSLATSRAALERRAVVVATERDEFLAGLKALASGQPAPGLVQGRVT
                     EGGLAFLFTGQGSQRLGMGRELYETYPVFADALDAVCVRLELPLMDVLFGTERDALDE
                     TGYTQPALFAVEVALFRLVESWGVRPDFLAGHSIGEIAAAHVAGVFSLDDACALVEAR
                     GRLMQALPTGGVMIAVQASEAEVLPLLTERVSIAAINGPQSVVIAGDEADAVALVESF
                     TGRKSKRLTVSHAFHSPHMDGMLADFRKVAEGLSYEAPRIPVVSNLTGALVTDEMGSA
                     DFWVRHVREAVRFLDGTRTLEALGVTTYVELGPDGVLSAMAQECVTGEDSVFVPVLRS
                     GRPEAESVTTALAQVHVRGIAVDWQAYFAGTGAQRVDLPTYAFQRRRYWLEEAPATAA
                     VEPLTGSLGAVDAQFWAAVDNADLSALTATLDIDVDADQPLSALLPALSAWRRQRQEQ
                     SVVDGWRYTVTWKPMADPAVARPTGTWLVVTPATSLVDLPAVSAALAAQGVDVREVAL
                     EAAELDRDGVAGRMREALAGDRADGVLSLLALAEHPHPAHPAAPTGLLLTGTLVQALG
                     DAGVDAPLWCLTTGAVATAPSDLIGSAAQAQVWGLGRVVALEHPERWGGLVDLPVPAD
                     ERALDRLLAVLAGAGDEDQIAVRSAGLLARRIGHAAPPAAGQHADSGTSGAGAAAGSA
                     WRPRGTVLVTGGTGALGGHVARWLAAHGAEHLVLLSRRGPQAPGADALVAEIAALGAG
                     ATAVACDVTDRTAVSELLAGLADGTYGPGLTAVFHTAGAGQFAPLDGTGPGEVAEVVA
                     AKVAGAAHLDELLGDTELDAFVLFSSIAGVWGSGGQSAYAAANAHLDALAQQRRARGL
                     TATSVAWGPWGEGGLVADDEAAEQLRRRGLPVMAPELSIAALQQALDGDETAVTVADV
                     DWDLFVPAFTAARPRPLITDLPEVRRALAAEQDGAATAAGEAAGLEAELRGMSGTEAE
                     GVVLNLVRTQVAVVLGHGGATAVEAARAFKELGFDSLTAVELRNRLSTATGLRLPASL
                     VFDYPTPAALAAHIRAELLGEDTTPELPALAEIDKLEFLLSSVPEDTTERARVTARLE
                     SLLSNWNRAERAVIGEDEEISIESASADDLFDIINNEFGKS"
     CDS             25902. .36188
                     /note="Orf10"
                     /codon_start=1

  Query Match             12.9%;  Score 129;  DB 1;  Length 161719;
  Best Local Similarity   49.9%;  Pred. No. 0.0011;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||    ||  |  |||| ||    |     || ||||||   |||||||| |||
Db     111857 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 111916

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||   |||||||| | |||| ||         |  | |  | ||  |  || | |
Db     111917 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 111976

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              |||  |  | | ||  |  ||  | |  || |     | || ||     || |||  ||
Db     111977 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 112036

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225
                          || | || |||||||| |  |||| |   |  |   |||||||| |
Db     112037 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 112096

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285
              ||  |||    ||| |    || || || |   | || ||| | ||  | |||  |  ||

Db     112097 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 112156

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345
              | || ||   |||  | |    |   ||  ||  ||  |  ||  |    |     |
Db     112157 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 112216

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
              || |     || ||| |  |  | | |||||| | |       ||     |  |||||
Db     112217 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 112276

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              |  |  | | | ||||| ||||  || |  |  |||      | || ||| || | ||
Db     112277 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 112333

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
                | |  || ||          || |||  |   ||| ||  || ||   ||     |
Db     112334 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 112393

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
                | | | |   || |   | |  |||| ||||    ||||   | ||||   |
Db     112394 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 112453

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
              |  |||  |   | ||||   |  | | |  | ||| | ||     | |    ||  |
Db     112454 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 112513

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
              |||  |    ||| ||| |   ||| || ||    |  || |  ||   | ||   || |
Db     112514 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 112573

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              ||| || || || || |||||||| ||||||||   |||  |||  | |||  |   |
Db     112574 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 112633

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
              | | || ||        |   || || ||| ||    ||   ||  | |||| |   |||
Db     112634 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 112691

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885
              |  |   | || || ||| |  ||| | ||  || | ||   | |   | |   | |||
Db     112692 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 112751

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945
              |  | || ||  |||    ||      ||||| |   |  | |||   | ||||
Db     112752 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 112810

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985
              || ||| |  ||  | || | ||| |||   ||| |||| |
Db     112811 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 112851


RESULT 14
CQ875911
LOCUS       CQ875911              164051 bp    DNA     linear   PAT 04-OCT-2004
DEFINITION  Sequence 18 from Patent WO2004065401.
ACCESSION   CQ875911
VERSION     CQ875911.1  GI:53789646
KEYWORDS    .
SOURCE      Streptomyces aizunensis
  ORGANISM  Streptomyces aizunensis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bachmann,B.O., Mcalpine,J.B., Zazopoulos,E. and Farnet,C.M.
  TITLE     Polyene polyketides, processes for their production and their useas
            a pharmaceutical
  JOURNAL   Patent: WO 2004065401-A 18 05-AUG-2004;
            Ecopia Biosciences Inc. (CA)
FEATURES             Location/Qualifiers
     source          1. .164051
                     /organism="Streptomyces aizunensis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:255848"
```

ORIGIN

Query Match 12.9%; Score 129; DB 6; Length 164051;
Best Local Similarity 49.9%; Pred. No. 0.0011;
Matches 499; Conservative 0; Mismatches 480; Indels 22; Gaps 6;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||    ||  |  |||| ||    |      || ||||||   |||||||| |||
Db     111857 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 111916

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||   |||||||| | |||| ||          |  | |  | ||  |  || | |
Db     111917 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 111976

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              |||  |  | | ||  |  ||  | |  || |     | || ||     || |||  ||
Db     111977 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 112036

Qy        181 C-----------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225
                        || | || |||||||| |  |||| |   |  |   |||||||| |
Db     112037 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 112096

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285
              ||  |||   |||  |   || || || |   | || ||| | ||  | |||  |  ||
Db     112097 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 112156

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345
              | || ||   |||  | |   |  ||  ||  ||  |  || |   |    |
Db     112157 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 112216

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
              || |     || ||| |  |  |  | ||||| | |      ||     |   |||||
Db     112217 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 112276

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              |  |  | | | |||||  ||||  || |  |  |||       | || ||| || | ||
Db     112277 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 112333

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
                | |  || ||          || |||  |   ||| ||  || ||   ||   |
Db     112334 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 112393

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
                | | |  |    ||  |  |||||  |||    ||||   | ||||   |
Db     112394 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 112453

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
              |  |||  |   | ||||   |  | | |  | |||  | ||      | |    ||  |
Db     112454 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 112513

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
              |||  |   |||  |||  |   |||  ||  ||     |  ||  |   ||  |||  || |
Db     112514 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 112573

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              ||| || || || || |||||||| ||||||||   |||  |||  | ||  |   |
Db     112574 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 112633

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
              | | || ||         |  || || ||| ||    ||   ||  | |||| |   |||
Db     112634 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 112691

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885
              |  |    | || || ||| |  ||| | ||  || | ||   | |   | |   | |||
Db     112692 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 112751

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945
              |  | || ||  |||     ||        ||||| |   |  | |||   | ||||
Db     112752 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 112810

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985
              || ||| |  ||  | || | ||| |||   ||| |||| |
Db     112811 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 112851
```

RESULT 15
CS073956
LOCUS CS073956 24081 bp DNA linear PAT 05-MAY-2005
DEFINITION Sequence 13 from Patent EP1524318.
ACCESSION CS073956
VERSION CS073956.1 GI:63090650
KEYWORDS .
SOURCE Streptomyces platensis
ORGANISM Streptomyces platensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Farnet,C.M., Staffa,A., Zazopoulos,E. and Yang,X.
TITLE Genes and proteins for the biosynthesis of polyketides
JOURNAL Patent: EP 1524318-A 13 20-APR-2005;
Ecopia Biosciences Inc. (CA)
FEATURES Location/Qualifiers
source 1. .24081
/organism="Streptomyces platensis"
/mol_type="unassigned DNA"
/db_xref="taxon:58346"
ORIGIN

Query Match 12.6%; Score 125.8; DB 6; Length 24081;
Best Local Similarity 50.5%; Pred. No. 0.0044;
Matches 419; Conservative 0; Mismatches 392; Indels 18; Gaps 4;

```
Qy        173 CGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGT 232
              | |||   | | | ||  |  |||    |||  || ||||| ||||| ||| ||||  ||
Db      18354 CTCGCCGTCCTTCGCCGTCGAGCTGGCCGGCCGCGAGTTCTTCCTCGACGACCACCGGGT 18413

Qy        233 GCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGA 292
              || |  |  | | | |||||||  | ||| |  |||||  ||||   ||| || |
Db      18414 GCGCAACGTGCCGGTGCTCCCCGGCGTGGCCTATCTGGAGCTGGCGTACGCGGCGGCCCG 18473

Qy        293 GATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCCGTCGT 352
              |   | |  ||      || | || || |  |||   |  ||||       ||| |  ||
Db      18474 GGCCGAGGGCGTCGACCCCGCCCACGCCCGCCTGCGCAACGTCGTCTGGTCGCGCCCCGC 18533

Qy        353 CGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGA 412
              |  | ||| ||||  || | |||||||  |||| ||||||  ||  | || |  |
Db      18534 ACGGATCACCGGGCCGACCGCGGTCGAGATCGCGCTGCGGCCGTGCGAGGACGACGCCTT 18593

Qy        413 CTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCG 472
              |  |  || |  | || | || |  |      ||  |      |    || ||||    ||
Db      18594 CACCTACGAGATCACCACGGCGGCCGACGGCGAACAGCCGGTGATCCACGGGCAAGGACG 18653

Qy        473 CCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCT 532
              | ||||   | | |   | |       | |  | ||| |||      |||  |   | |
Db      18654 CATCGAGCGGTGCGGGACGCC------GTCACCCGCGCGCCTGGACATCGCCGCGCTGCG 18707

Qy        533 CTCACGTTGCTCGAGGGTCACCGGCGGA---GACGCGTGCTACCGCCGCCTCCAGCAGCT 589
              | | |  |||  |  |  |||    | |   ||||  |||||||| | | || | |   |
Db      18708 CGCCCAGTGCGAGGTGCGCACTCTGGAACACGACGACTGCTACCGGCTCTTCGACCGCAT 18767

Qy        590 CGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGA 649
              ||  |   | |||||      ||||||  |   |||    | | | |   ||
Db      18768 GGGCATCGGCTACGGCCCGGCCATGCGGGGCATCCGGCGGATCCACGTCGGCGCCGGGCT 18827

Qy        650 GGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCT 709
              |||  ||||      |  |||| ||| ||      |  |  ||||  | ||  ||  | |||
Db      18828 CGCCGTCGCACGCCTGAGCCTGCCGCAG---GCCGCCCGGGACGGCGCCGGCTGGGACCT 18884

Qy        710 CCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGC 769
              || ||  |  |||||||||| ||| | ||   |||  | || |||       | ||||
Db      18885 GCACCCGTCGATGCTCGACGCCGCCGTACAGGCCACCTTGGGCCTGTCACTGGCCGAGGA 18944

Qy        770 GGCCCGAGCC------CCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCC 823
              ||    ||       || || |||| | | |     |||   |||||||   ||
Db      18945 CACCGACACCGTGGCGCCGGCACTGCCCTTCGTCCTGGAGGAGGTGCAGCTGCTCGCGCC 19004
```

```
Qy        824 GCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGC 883
              |||||  |    |  |||   |     ||  |  | |        |||     ||
Db      19005 CAGCCCGGCCGGCGGGTGGGCCGTGGTGCGACCCGCAGCGGGCGACGGCGGCGGAGCCGT 19064

Qy        884 CAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGA 943
              |   |    || ||   |||  |||   || ||||  ||||  | |     |||  |
Db      19065 ACGCCGCATCGACATCGAACTGTGCGACGACGACGGCGAGGTGTGCGTACGCCTGCTCGG 19124

Qy        944 TTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992
              |||||||  ||| ||  | | || |  | ||  ||| ||  ||   ||
Db      19125 GTTCACCGCACGCGTCCTGGCCGCCGGTGACGACCCCGCCGGCGGAGAG 19173
```

Search completed: April 28, 2006, 02:39:35
Job time : 5544.08 secs

```
                /note="similar to herpes virus membrane glycoprotein"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20165.1"
                /db_xref="GI:13432084"
                /translation="MPPRVRHQQIQHGRGRRPLLHRRRHGQRLRGDPGGHRDTAPPAR
                RAAACRARHARACARATNRPHSAEWYRTRPAGTRSEQPSRPHTGRDAARRDRFRSPPA
                RPWHCGEQYRRSRAACRAVGGTTTVQPSRAHSRATGTRPTATSTASPATSSVIGSVAA
                SSARSRPVSPPYTPSGAYGTVTGCPSTTSTTGTIASAAHSARTGQAAHTRRAAPTTTA
                PSTPAASSWPSSPASSSASAASANRAGQSAHGRGAPSTTARNTGTDHDGNASNSYGST
                ARISAWTGIVGAVIATRPCRSRSRPSAAAPHPGPAATSATTGHPRHRGQRGAAPGAVG
                HAASQVDLRCRSAQAPGRR"
CDS             1128. .1562
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20146.1"
                /db_xref="GI:13432065"
                /translation="MRRYLDPDGLVHGLPTYPYRYAPDGLYTRSQLREMGLRPIADPV
                AQVMWDGSARHRRSGEPVRTAALYDITQTLLRDTPSLTRLAQLAIARACRRICPACDR
                MWPYVIPAAAGMCLDCQTLTAAGVPVGSSTDVFDPTAGRAAA"
CDS             1559. .2929
                /note="similar to Streptomyces coelicolor putative
                secreted protein and Enterobacter plasmid R751 KfraA
                protein"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20147.1"
                /db_xref="GI:13432066"
                /translation="MSERIERAVAAVAPDALGATAFAVSFTHVQHVLETHGAHGWTSW
                ALSVALEALAITAVLELRRRRRLAQSVLPPVGVLLTGVALVVACNLESATAKGWGTVV
                ALLGPLLYLLAWSLVATRGRSGPDPRAERDAALAELAEAAAARDWAAGVVETVRGELA
                ALSAERDQAHAGLAAAEAARAELAAEVAAVRAAQEAAVTEAAAAERDAMAAAVTVACE
                TRDQALAERDAIRVQAEDAARQATEARAAALAAERDAAGARQAAARATADLTEVRSER
                DAALGETRAARADRERVEQTARETAAALAAAEAERAGLAGELVATSRDLSQARADLEA
                AASREVSQTAEQTGAGKGAGKGSGLAEWTADATPQELAALVAVLETMQVRRTGKGLTY
                AAARGVLGVRHAVVRTAVDTARGSLRAGTPIFPGGFPGDDIEDAEIVPMPSPRELESA
                AAGGRR"
CDS             2926. .3156
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20148.1"
                /db_xref="GI:13432067"
                /translation="MSAGDRRGQVLAALAASPLIDPAASVTQLLRLADELLAAAVPAV
                EATAMDERLRAEMAASTGTAWTTETTGRRWAA"
CDS             3153. .4727
                /note="similar to TrbI transfer factor from plasmid RK2"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20149.1"
                /db_xref="GI:13432068"
                /translation="MSGSAAETQAPEQISGGWTPVVEETPAPPAPPVRRRGVPAGPVI
                IAAGEAGVMAGSSLYAAAGVPGLIAAGAVAGTAAAARTVRVAQRRAASRASTSGGGAG
                RGRSGSGPVPAALRRLAGAGRRRGAGGGPGTGKGAGTGTGTGRRRRGAGTGRRPGSGR
                AGGAAGRTAGGRRAGGGPLTRAARGTGGQAARAARRSGAAAVRAARLRGAGMVRAARI
                GRSAAGRLAAAVRQARSPWAAARAARRGRYRMVAGQRRRRGGGLGRLVRTGLGLLAAL
                AAVGWWLGHRGWRLTRAGWRRLRGPRPGPDGRPGTDLIGPLAQLRTPPIGGPAAPVPP
                PAGGGAMDPQAILAGARAALAGVAFATFGRLPEMPEMPPIPGGPVASLMPGHDRGDAA
                MAARDRDGRSLAGGSGGVFLLPEQAETMMQTAAAYAPPSGPQIGRDMAQVPQAVEHIA
                AAIAGLSRLAADELPVHETIRELLDSAAAQIVAAAGTLANLGPLFELVHQTDLHRLRQ
                PRPGEGLWDSHHAGDA"
CDS             4729. .6942
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20150.1"
                /db_xref="GI:13432069"
                /translation="MAADSTPDSSGQQTQTAGRGRRWEWEWTAAHGPLAGTVNAAGAA
                MSAAMVGDLAGVNPLWAAGAGLAGAGASALAGAHSGLSRAGIAYRVACWTAAGGWTWH
                AIGSGAWSLDSAGALVVGGVAGGLLAPVFARHERQESARRRRMLLATARARTAEEWVD
                RIERVCRIRGVQIVSVVDWESGAGYSMDVDLPRGGASWRTLARFSEDLANDADLPAGA
                GVEVLPGTTRRSAVLEISTVDGFAEDLLFGDDYSELTINNPLMIGRYRNTQPAEAHLR
                ETYALVVGEQGSGKTNQLYVLTGQLARCTDVIVCHIDPNGGGLPRPWVMPWVEGQATR
                PAVDWVAYEPVEAERMTGALLRGMTRRKGAYHGRMRAANDDKLPIDSSVPAIVVIVDE
                AAEIMAPDAPVEWRPTRDNLRAIQRQGRAMGMLIIFAGLRATVEVLGTTTVKKLTQLR
                IGMGVTDQVERAILFGQASTMDGDQVLGRGYGHLSYRGEQPRAFKGLRIKPETITALT
                RAVAGRRPAPDAPTLEAFGADWAGRWERTLPGLVAASDTDGDGPASPEDLVAFAAGLA
                AQANGEDPNAAMDQARAAGPEAAQDLLDDVMREGPADADTANARAADAVEKARRLLAE
                QRAWRELVDGWDATAPDPQSTPPADGGAGGEEIDTDAARARMIEMLYAAGPKGMAAGV
                LWKALAAEGIDRTRQAVGDWLRLAIVAGQVVQLRERGPYYHADHAPKGPDGDDGQ"
misc_feature    6942. .7010
                /note="similar to GenBank Accession Number AF014839"
CDS             7117. .7581
                /note="similar to Pseudomonas hypothetical protein"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20151.1"
                /db_xref="GI:13432070"
                /translation="MTPHKIAAPPPATRLTSTLIVGRVGTGKTVETVARTRAHLDAGG
                LVWLAGDPLPAGGVDHRRLDWHERGADGIRSLIEVACRLAAGRLRLDAAVGPLLLVAD
                PWPTGPAAGRLTWLAEHGPTGVTPVGVVATALRPPPGAVWGERIDLPAALPL"
CDS             7663. .8043
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20152.1"
                /db_xref="GI:13432071"
                /translation="MPIRVRAQRRLGPLRATITERGLSSVTLRVGRTSWRLWSRTGGR
                RGLTSVNLPGPLSWRPAARPRTAGQRAAAAGRSRRRWLLLTVLALVILAGAAAHGVDG
                RPLAALAVTAAVCGAGYLHRRRRP"
CDS             7773. .8471
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20153.1"
                /db_xref="GI:13432072"
                /translation="MVAYRRPPRADERQPARPAVVAAGGPAADRRPACRGGRPVPPPL
                AAAHRPRPGDPRRGRGARRRRPAPGRPRRDRRRLRRRLPAPPPASLTHTVRPPLPGGS
                VTIPTTIPALLRAAAGEITTRGWWQHDYLGDDGEVCLIGALRAAAGAHPLREPEEPDA
                HDLLGAALADLALRIPADVDEDPDPEVGLGLFDTGAGAAIADWNDADGRTADQVTTLL
                HTVAADLDRPAAAA"
CDS             9046. .10236
                /note="similar to Mycobacterium tuberculii hypothetical
                protein and Bacillus subtilis SpoJ chromosome partitioning
                protein"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK20154.1"
                /db_xref="GI:13432073"
                /translation="MSVSSVSVRPGSTRLLIRRSWSACCGGRRPARCCGSGGAVGARP
                GCAAVSCRTPSPPTTRWTSCGPRCVGCCGPALRPGPLPAWPSADAAAPPAGDAPSFSV
                PGSGGASPRLGAPCTVSVDVYGAPDYGPIMPAVRITIAAHKGGVGKSATTAALAGALA
                LTGHRVLAVDTDPQGALGALLGVPSPQPPGLYEVLTGGVPATAAVVRSSTPGLLLLPA
                TRDLAAADLELPRRAGWRRALAAALATIPAGAADVILIDSPPGLGVLPVLALGAATHV
                AVVTELEHLSIRALPDALDTVRQVVTTAGGSPRLLGIVPNSVDLRTLHQRDAYGRLRE
                LYEPWILPGLPARVSVKEAGLAGLPISLYDPHGDITVAVTALAQEVARRALVPDAPTP
                AQHE"
CDS             10714. .11430
                /note="similar to Streptomyces coelicolor hypothetical
                protein"
```

```
                     /gene="SCO3285"
                     /note="SCE15.02c, large glycine/alanine rich protein, len:
                     1772aa"
                     /codon_start=1
                     /transl_table=11
                     /product="putative large glycine/alanine rich protein"
                     /protein_id="CAB41271.1"
                     /db_xref="GI:4678899"
                     /db_xref="UniProt/TrEMBL:Q9X877"
                     /translation="MATAPEKAATGTHDTGDTADAGAYEVLRDRLAAQTADLARRAGL
                     LNARRVEEFGSARLELASTERLRTEHPGVPCDLAAVGDALLLGSTGRPGHRTGTAAVA
                     DVFTLYDRDLNPLPESAVPGLLDDPAFVREFADLHRYYRQARLLRLRPVGGRLLAVFR
                     TGEKAADIRVLRWELTEDGRAAFLDARGERDDVFPPSHDFEWTAATREDHVLGRHPHV
                     SVRGEFHVSAVGGALTVKLEDDTETAGGVYSEPVDEPLQSLADADIAHARVGALILLR
                     VRPYKEDTDRHLVFNTLTKSVVRLDGIGAACRRLPDDQGIVFPGGYCLATGVHKTYEL
                     DAAGLEFEREVRSPNGEDVLYAFHARGESRGLLLSYNTIRKEVANPLPCRGWALAEDG
                     TFTVLRADGDEPAQVHPVQLWHSPYVSDTHAAAAPAGSGPLARVGNADLVRGISACLS
                     VAGAVGDGITTAEGYRALAASCVRAADGHHWLGEADLGDLAGALAAVRETAEQVLAEF
                     ETVRDLTRRAAEARDEAAERIASVVRRLRGEAPKEAAAWVRGLTELRHAHGHLLTVKE
                     MRYADAPGIDALAGEAEASLAELGRRAVAFLAREDAFDAQRADVEALVADAEAIATVA
                     EAGPVAARLDELADGLRTVTDVVAELDVGDATVRTALLERVAAVLGGVNRARATLDAR
                     RRALLDREGRAEFTAETALLGQAVTAALAAADTPERCDDQLARLLARLEDLESRFAEF
                     DGFLAELADKRTEIYDALAARKQALSDTRARRAEQLAASAARIMETIARRCATLADAD
                     AVSTYFASDPMPAKVRRTAEELRALGDSVRAEELDGRLKSARQEASRALRDRTDLYAD
                     DGRTLRLGAHRFAVNTQPLDLTLVPDGDGLAFALTGTDYRSPVTDPDFAATRAHWDRT
                     LPSESPGVYRAEHLAARLLRQHGASALAAADDLPALVREAAQEAYDEGYERGVHDHDA
                     TLVLTALLPLYEKAGTLVHEPAARAAAQLFWAHGTTPRAREAWTRRARSLARARDTFG
                     LSAAIAALEEELAGAVGTWPGPGEMVAGAGPVPGGAAAGAPAEDGRAGVPPAATGGAV
                     GSLPGASGPPSGAGAPLGAAGSAAGTAGSGAGAPGSGVGAAVSGAAGDAGRAAGVPRA
                     AGGTARSAEGFAGAAGAVGSATGTSGPPAGAPGSGAEHARAAAAYLFHELTTEPEGFV
                     LGAGTRTLLEKFRRTVGTPAYDEDLAALDDLAARGQLAEAWISSYAAAGGTGLTPGDL
                     AEAVAAELCPDLPRYDGDVPPTATAEGLLGTHPRITGGRLQLRLDEFLARTARFAAHD
                     VPEFRAYQRRRTALVGAERARLRLDDHRPRVMSAFVRNRLVDEVYLPLVGDSLAKQLG
                     ATGDGKRTDTGGLLLLLSPPGYGKTTLVEYVAERLGLMLVKVGGPALGHGVTSLDPAD
                     APNATARQEVEKINFALASANNTLLYLDDIQHTSPELLQKFIPLCDATRRVDGVWNGA
                     PRTYDLRGKRFAVCMAGNPYTESGARFQVPDMLANRADVWNLGDVLTGKKEAFALSFL
                     ENALTANPVLAPLAGRDRADLELLVRLATDDPTARADRLDHAYPPAELERILSVLRHL
                     VAARETVLAVNAAYIASAARSDETRTEPAFRLQGSYRNMNKIAQRVRPVMNDAERAAV
                     LDDHYTAEAQTLTHGAEANLLKLAELRGTLTPEQADRWAEVRTAHVRARTLGGPDDDP
                     LTRAVAALGLLADRVAAVESAITRAADPRHLLANPHARHAAGGTER"
     gene            complement(9259. .11292)
                     /gene="SCO3286"
                     /note="synonym: SCE15.03c"
     CDS             complement(9259. .11292)
                     /gene="SCO3286"
                     /note="SCE15.03, putative secreted protein, len: 677 aa;
                     weak similarity to SW:YQIK_ECOLI hypothetical protein from
                     Escherichia coli (553 aa) fasta scores; opt: 542, z-score:
                     501.2, E(): 1.3e-20, (27.0% identity in 504 aa overlap).
                     Contains possible N-terminal region signal peptide
                     sequence also contains a highly hidropholic central part
                     of the protein"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secreted protein"
                     /protein_id="CAB41272.1"

  Query Match             10.9%;  Score 97.8;  DB 1;  Length 293050;
  Best Local Similarity   47.0%;  Pred. No. 5.5;
  Matches  380;  Conservative    0;  Mismatches  417;  Indels   12;  Gaps    2;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              || | | ||||    |  |  |     | ||| | |||| ||  | | || |||
Db       7812 CGTCGGCGACGGCATCACCACCGCCGAGGGCTACCGGGCGCTCGCCGCCTCCTGCGTGCG 7753

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
                | ||    |||| | | |  | || |||| | |  ||     | |  |    |||
Db       7752 AGCGGCCGACGGCCACCACTGGCTGGGCGAGGCGGACCTGGGCGACCTGGCCGGCGCGCT 7693

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              | |||||||   |   |  | |  | ||| |  |  |  | |     | | | |||
Db       7692 GGCCGCGGTCCGGGAGACCGCCGAGCAGGTGCTGGCCGAGTTCGAGACGGTGCGCGACCT 7633

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              | |||  |||||| |  |  |  |    |  | || ||  ||| ||    || ||  ||
Db       7632 GACCCGGCGGGCGGCCGAGGCCCGCGACGAGGCCGCCGAGCGGATCGCGTCGGTGGTCCG 7573

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
                |  |||| | | |    | |  |  ||||  |  | | |   ||  |
Db       7572 GCGGCTGCGCGGCGAGGCGCCCAAGGAGGCCGCCGCCTGGGTACGGGGCCTGACCGAGCT 7513

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
              |||  |   |  |  || || |   ||| | |||||||  |  | || |    | ||
Db       7512 GCGGCACGCCCACGGGCACCTGCTGACCGTCAAGGAGATGCGGTACGCCGACGCCCCCGG 7453

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              | |||||  |||||| |  || | ||  |    |  | ||||  | |  |||||||||
Db       7452 CATCGACGCCCTCGCCGGGGAGGCCGAGGCGTCCCTCGCCGAGCTGGGC---CGGCGGGC 7396

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              | |||       ||  |  | |  | ||| |||||| | ||     |    |
Db       7395 CGTGGCCTTCCTCGCCCGCGAGGACGCCTTCGACGCCCAGCGCGCGGACGTGGAAGCGCT 7336

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||    | ||   |    | |||    | || | |  ||  | | ||  | |||   ||
Db       7335 CGTCGCCGACGCGGAGGCCATCGCCACCGTCGCCGAGGCCGGTCCGGTCGCGGCCCGGCT 7276

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
                |||  |  | |||  | |  |         ||| |||| |||          | ||  ||
Db       7275 CGACGAACTCGCCGACGGGCTGCGCACGGTCACCGACGTCGTCGCCGAACTCGACGTGGG 7216

Qy        602 CAGCAAGATCGGCGACACGCGCTT---------CGTCGGCCACCTCCAGCGGCTCCGCCG 652
              |  |   | ||  |  |||   | |         |||| |  |  ||| | |   |  |
Db       7215 CGACGCCACCGTCCGCACCGCCCTGCTGGAACGCGTCGCCGCCGTCCTGGGCGGCGTCAA 7156

Qy        653 CGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGC 712
              |    || ||   |    || |||||| |          ||| |  ||    || |||
Db       7155 CCGCGCCCGCGCCACCCTCGACGCCCGCCGCCGCGCCCTCCTCGACCGCGAGGGCCGCGC 7096

Qy        713 GGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGA 772
              |  |   || ||  |||| |  |||| |  |||| |    |    |   ||| | ||
Db       7095 CGAGTTCACCGCCGAGACGGCCCTGCTCGGCCAGGCCGTCACCGCCGCGCTCGCCGCCGC 7036

Qy        773 CGACCCCGCCGAGCTCCGCCTCGTCCTGC 801
              |||| || |||||| | ||  || || ||
Db       7035 CGACACCCCCGAGCGCTGCGACGACCAGC 7007


RESULT 5
AY027524/c
LOCUS       AY027524               22437 bp    DNA     circular BCT 18-APR-2002
DEFINITION  Frankia sp. CpI1 plasmid pFQ12, complete plasmid sequence.
ACCESSION   AY027524
VERSION     AY027524.1  GI:13432064
KEYWORDS    .
SOURCE      Frankia sp. CpI1
  ORGANISM  Frankia sp. CpI1
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Frankineae; Frankiaceae; Frankia.
REFERENCE   1  (bases 1 to 22437)
  AUTHORS   John,T.R., Rice,J.M. and Johnson,J.D.
  TITLE     Analysis of pFQ12, a 22.4-kb Frankia plasmid
  JOURNAL   Can. J. Microbiol. 47 (7), 608-617 (2001)
   PUBMED   11547880
REFERENCE   2  (bases 1 to 22437)
  AUTHORS   John,T.R., Rice,J.M. and Johnson,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2001) Molecular Biology, University of Wyoming,
            Laramie, WY 82071-3944, USA
FEATURES             Location/Qualifiers
     source          1. .22437
                     /organism="Frankia sp. CpI1"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1856"
                     /plasmid="pFQ12"
     CDS             complement(join(22201. .22437,1. .825))
```

```
RESULT 4
SCO939116/c
LOCUS       SCO939116             293050 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 13/29.
ACCESSION   AL939116 AL035636 AL049573 AL049628 AL049630 AL049645 AL049661
            AL049707 AL049763 AL049819 AL049841 AL079345 AL133236 AL133252
            AL645882
VERSION     AL939116.1  GI:24413781
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
   PUBMED   12000953
REFERENCE   2  (bases 1 to 293050)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 26, 2002 this sequence version replaced
            gi:4539555, gi:4582361, gi:4585581, gi:4585828, gi:4678647,
            gi:4678625, gi:4678897, gi:4757084, gi:4803679, gi:4808332,
            gi:5123647, gi:6469451, gi:6491805.
FEATURES             Location/Qualifiers
     source          1. .293050
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     gene            57. .713
                     /gene="SCO3282"
                     /note="synonym: SCE39.32"
     CDS             57. .713
                     /gene="SCO3282"
                     /note="SCE39.32, possible precorrin-8X methylmutase, len:
                     218 aa; similar to many e.g. SW:COBH_PSEDE Precorrin-8X
                     methylmutase from Pseudomonas denitrificans (210 aa) fasta
                     scores; opt: 868, z-score: 1000.1, E(): 0, (64.2% identity
                     in 204 aa overlap) and SW:COBH_MYCTU precorrin-8X
                     methylmutase from Mycobacterium tuberculosis (208 aa)
                     fasta scores; opt: 852, z-score: 982.0, E(): 0, (64.1%
                     identity in 206 aa overlap)."
                     /codon_start=1
                     /transl_table=11
                     /product="putative precorrin-8X methylmutase"
                     /protein_id="CAB40340.1"
                     /db_xref="GI:4582393"
                     /db_xref="GOA:Q9X8F5"
                     /db_xref="InterPro:IPR003722"
                     /db_xref="UniProt/TrEMBL:Q9X8F5"
                     /translation="MTPYDSKNEYEYEYEKDGAAIYRQSFATIRAEADLSGLPADVGR
                     VAVRMIHACGMVDLVRDLSWSPGVVSRAREALRAGAPVLCDVRMVASGVTRARLPADN
                     DVVCTLSDPAVPELAARLGTTRSAAALELWRDRLDGAVVAVGNAPTALFRLLEMIGEG
                     APRPAAVIGVPVGFVGAAESKEALAAHPSGVEHLVVRGRRGGSALAAAAVNAIASEEE
                     "
     gene            complement(745. .1512)
                     /gene="SCO3283"
                     /note="synonym: SCE39.33c"
     CDS             complement(745. .1512)
                     /gene="SCO3283"
                     /note="SCE39.33c, possible precorrin-6X reductase, len:
                     255 aa; similar to many e.g. SW:COBK_RHOSO precorrin-6X
                     reductase from Rhodococcus sp. (248 aa) fasta scores; opt:
                     873, z-score: 978.5, E(): 0, (55.1% identity in 247 aa
                     overlap) and SW:COBK_MYCTU precorrin-6X reductase from
                     Mycobacterium tuberculosis (244 aa) fasta scores; opt:
                     690, z-score: 775.5, E(): 0, (50.4% identity in 244 aa
                     overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative precorrin-6X reductase"
                     /protein_id="CAB40341.1"
                     /db_xref="GI:4582394"
                     /db_xref="GOA:Q9X8F6"
                     /db_xref="InterPro:IPR003723"
                     /db_xref="UniProt/TrEMBL:Q9X8F6"
                     /translation="MHVLILGGTTEARRLAELLAAGQPAGPRVTTSLAGRVSAPRTPP
                     GEVRVGGFGGAEGLAAWLREHAVGLLVDATHPFAGTITRNAAHAAATAGVPLLVLRRP
                     GWRPVEGDVWHEAGSLAEAAALLPALGRRVFLTTGRTGLAVFAPLADPWFLVRSVDAP
                     EGPRPPRTEVLLDRGPFTLDGERELLRRHRIDVLVTKDSGGAATAPKLTAAREAGLPV
                     VVVRRPPVPEGVRVVAEPGQAARWVAEARARRSPRRP"
gene                 1683. .3818
                     /gene="SCO3284"
                     /note="synonyms: SCE15.01, SCE39.34"
CDS                  1683. .3818
                     /gene="SCO3284"
                     /note="SCE15.01, partial CDS, conserved hypothetical
                     protein, len: >436aa; similar to C-terminal region of
                     TR:CAA22732 (EMBL:AL035161) another conserved hypothetical
                     protein from Streptomyces coelicolor (860 aa) fasta
                     scores; opt: 922, z-score: 1011.5, E(): 0, (39.5% identity
                     in 425 aa overlap).
                     SCE39.34, unknown, len: >307aa; similar to TR:E1370406
                     (EMBL:AL035161) conserved hypothetical protein from
                     Streptomyces coelicolor (860 aa) fasta scores; opt: 254,
                     z-score: 277.0, E(): 4e-08, (29.2% identity in 277 aa
                     overlap) TR:O86875 (EMBL:U22894) hypothetical protein from
                     Streptomyces lividans (547 aa) fasta scores; opt: 226,
                     z-score: 249.7, E(): 1.3e-06, (33.0% identity in 197 aa
                     overlap)."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD55311.1"
                     /db_xref="GI:24413782"
                     /db_xref="GOA:Q8CJX1"
                     /db_xref="InterPro:IPR000014"
                     /db_xref="InterPro:IPR000700"
                     /db_xref="InterPro:IPR001610"
                     /db_xref="InterPro:IPR001932"
                     /db_xref="InterPro:IPR002197"
                     /db_xref="InterPro:IPR003018"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="InterPro:IPR010822"
                     /db_xref="UniProt/TrEMBL:Q8CJX1"
                     /translation="MRKARGKAAKGGRAVTGDPIGDAAVLSALFANSPQGLFVFDSEQ
                     KVTRYNPAGRGVRGLPEEDVVGQGIADFVPGFEAVELQGLIDEVLATGEALRGRLVRG
                     RAPSGPHRTLAVEVSLFPLRDLGDGKPGLVGVVEDVTERQAAADRLAVLSEVHATVGS
                     TLDVRTTGDELVAALVPAFADAATVDLLDDGVVAAERAAGPPPTEVPLRRVAFAPPAA
                     EVSRKEGDSRPLPFPTPYTQVLNDMRARLVRVSSDDPWLSADPEALEPLVRTGVHSMI
                     VAPLLARDTVLGLLTLYRHRTDPFEEADLDVARQAAFTASVHLDNAHSYRREHTVAAA
                     LQRRLQPGAIPLLSAVETAHVYLPESAGGDWFDVVPLSGTRVALVVGDVTGHGIEAAA
                     TMGQLRLAVRTLALRDLETDELLTHLDEVASQLADASGPGGNTPVATCAVTVYNPVSR
                     HCTMVRAGHPAPVVIDPDGVPRAVDVPQGPPLGTGGGHVFTPAVLELAPGTLLALHTN
                     GLTRGNGDEGEAARRRLEHILASTTRPLQELCDTAVYHMAPSRHDDAVLLLARTRALP
                     PEHVADWTLPADASVVGTARRLVDRQLASWGLDEAAYTTGLVVSELVTNAIRYGKGPV
                     RLRLIRDRGRLLTEVTDANSASPHLRRARENDEGGRGLFIVMRLSTHWGVRHSRHDKT
                     IWSEQRLDAGPPDTSGLLDAFDVTDAAGL"
gene                 complement(3855. .9173)
                     /gene="SCO3285"
                     /note="synonym: SCE15.02c"
CDS                  complement(3855. .9173)
```

```
Qy          5 CCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGG 64
              |||||| |||||| || |  |  |  |  |  |  ||   ||||||| || |
Db     214084 CTGGGAGGAGGACGCCCTCGACGCCGAGGAGCTGGCCGACACCGTCCTCGGCCTGGTCCG 214143

Qy         65 GGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGA 124
               |  |||   |||  ||||  | |||||   | ||    | ||  || |||  ||||
Db     214144 CGACGCCGGACTCGACCCGGGCGACGAGCCCTGGCTGGGCGCCCTCGCCCTGCCCGACGA 214203

Qy        125 CGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTC 184
              |  | ||    ||| |    |    |    ||  |||    ||   ||  || |
Db     214204 GGACGGCGAGCTCTCCCCGGCGGGCGAACTCGTCTTCCCCGGCGGCCCGTTCGCCCGGGT 214263

Qy        185 CCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCC---A 241
              | | |||| |  |     | |  |  || |  |  |    |  | | |   |
Db     214264 CATGCGGGAGGACGAACTGGCCGCCGTGGACGCGGAACTCGCCGAGAAGTGGGGGCCGGA 214323

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              |   |  || |  ||| | || |   |  |  |   | || |  |||||||
Db     214324 CCCGCTGGCCGCCTGCGGCGTGCTCGTCACCTTCGCCCTCGTCCGCGCCACCGACGTCGT 214383

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
              |  || | |  | | |    |  ||| | |||  ||| |||| | |    | |
Db     214384 CCTGGACCCCGACGAACTGGAGCCCCGCGAAGGGGACTTCGCCGAGCCCGACGACGCGGG 214443

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              ||   | ||  || || ||     ||| | ||   ||| ||  | | ||       |
Db     214444 ACTGCTCGACGCCGTGGACGTGTGGAGCGAGGACGTCCTCGACCGCTTCCCGGACACCCC 214503

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              || |  ||    |||||      |  |  ||    | |||||| |   |   |   ||
Db     214504 CGTCCCGCCGGTCGCCACCGAGATCGTCGCCGTGCGCGACCTCGACCTCGTGGACGACGA 214563

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              | |    | |   ||||||  |  ||| | || | |  ||| ||| || | |  ||  |
Db     214564 CCACTGGCCCGAGGCCCTCGCCCTGCTGTCCCGGCCGCCGCTGCGCGACGCGCTGGTCCA 214623

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | |||  || |||  | || |      |  |  ||| |||| ||| |   |  | |  |
Db     214624 GCCCGTGCGCGTCCTGCTGCCCGACGGCACCCACGAGGTCGTACGGCCCTACACCGCGTG 214683

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              |   |  ||| || ||    ||| ||| | || ||      ||  |||| |
Db     214684 GTGGCTGCGCGGGCACCCGGTGCTCGGCGGACGCCGCCCGGCCGGCCTGCGCGCGGCGGG 214743

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              |||      ||  |  ||  | || | |      | |||||    |  |   |   |
Db     214744 CGGCGACCCGCTGCTGCGGGGCCTGTACGACGAGGCCGACGCGACCGGCTTCGAGGACGA 214803

Qy        722 CACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGC 781
              |  ||| || |||| ||   |  | | ||  || |    ||||||| || |
Db     214804 GCAGGTACTGCGGGCCCTCGGCGTCCGCACCTCGGTCGCCGCCCTCCTCGACGAGCCGGG 214863

Qy        782 CGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGG 841
              ||       |  ||||| || |  ||| | |  | |    ||    |  |    |
Db     214864 CGGCGCGGCCGAACTCCTGGACCGCCTCGCCGACCCGGACCGCCCGGTCACGGCCGCCCA 214923

Qy        842 CAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCGGA 901
              | | | || |    |  | |  ||||  || | |||||  ||    |    ||||
Db     214924 ACTGCACGCCCTCTACGGCGCGCTGGCCGAGCTGGACCCCGAGCGGGTCACCCTGCCGGA 214983


RESULT 3
CP000081_01
WPCOMMENT
  Sequence split into 21 fragments  LOCUS CP000081 Accession CP000081
     Fragment Name      Begin     End
     CP000081_00            1    110000
     CP000081_01       100001    210000
     CP000081_02       200001    310000
     CP000081_03       300001    410000
     CP000081_04       400001    510000
     CP000081_05       500001    610000
     CP000081_06       600001    710000
     CP000081_07       700001    810000
     CP000081_08       800001    910000
     CP000081_09       900001   1010000
     CP000081_10      1000001   1110000
     CP000081_11      1100001   1210000
     CP000081_12      1200001   1310000
     CP000081_13      1300001   1410000
     CP000081_14      1400001   1510000
     CP000081_15      1500001   1610000
     CP000081_16      1600001   1710000
     CP000081_17      1700001   1810000
     CP000081_18      1800001   1910000
     CP000081_19      1900001   2010000
     CP000081_20      2000001   2090491
Continuation (2 of 21) of CP000081 from base 100001 (CP000081 Leishmania major chromosom

  Query Match           10.9%;  Score 98.6;  DB 2;  Length 110000;
  Best Local Similarity  45.4%;  Pred. No. 7.1;
  Matches  353;  Conservative    0;  Mismatches  424;  Indels    0;  Gaps    0;

Qy         34 TCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGC 93
              || |||   ||    ||    ||   ||  |   ||  || ||     || | | |||
Db      74215 TCTGCGCCGTCCAGCAGCAGCACCGCGCCGTCGGCGTCCTCGTCGTCTGCGCCGTCCAGC 74274

Qy         94 GGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGC 153
              | ||  ||  ||| |           | || ||    || ||  ||  ||  ||| |
Db      74275 AGCAGCAGCTCCGCGCCGTTGGCGTCCTCGTCGTCTGCGCCGTCCAGCAGCAGCAGCTCC 74334

Qy        154 CGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCG 213
                  |   | | || ||  | |   ||  | |  || |  ||     |   |
Db      74335 GCGCCGTCGGCGTCTTCGTCGTCTGCGCCGTCCAGCAGCAGCACCGCGCCATCGGCGTCT 74394

Qy        214 GCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGC 273
              || || | |   | | |  || | | || |||||||  |   |  | |||  |
Db      74395 TCGTCGTCTGCGCCGTCCAGCAGCAGCTCCGCGCCATCGGCGTCCTCGTCCTCTGCGCCG 74454

Qy        274 GCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACG 333
              |||   | | |||  || |  || |   | ||    ||| |     | | |      |
Db      74455 TCCAGCAGCAGCTCCGCGCCATCGGCGTCCTCGTCGTCTGCGCCGTCCAGCAGCAGCTCC 74514

Qy        334 AGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACG 393
              |  |||||||    ||  |||| |  | || | || |     |  || || || ||
Db      74515 GCGCCATCGGCGAACTCGTCGTCTGCGCCGTCCAGCAGCAGCAGCTCCGCGCCGTCGGCG 74574

Qy        394 GCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCG 453
              |   | ||  ||||||  | || |  | ||  |||       | ||||| |  |  ||
Db      74575 TCCTCGTCGTCTGCGCCGTCCAGCAGCAGCTCCGCGCCGTCGGCGTCCTCGTCGTCTGCG 74634

Qy        454 GCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCC 513
              || ||| |   | ||    |  |  | || |||  | ||  | | || |||      |
Db      74635 CCGTCCAGCAGCAGCTCCGCGCCGTCGGCGTCCTCGTCGTCTGCGCCGTCCAGCAGCAGC 74694

Qy        514 CGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCA 573
                  |||   | |    |  |      |||  | |    || |       ||  |
Db      74695 TCCGCGCCGTCGGCGTCATCGTCGTCTGCGCCGTCCAGCAGCAGCTCCGCGCCGTCGGCG 74754

Qy        574 TCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCC 633
              ||  | ||||  |  |   |||    |       ||  ||  |   | | |||||  ||
Db      74755 TCCTCGTCGTCTGCGCCGTCCAGCAGCAGCTCCGCGCCGTCGGCGTCCTCGTCGTCTGCG 74814

Qy        634 ACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGA 693
              | ||||||  ||    |||||   ||  ||              |  |  ||  |  ||  |
Db      74815 CCGTCCAGCAGCAGTTCCGCGCCGTCGGCGTCCTCGTCGTCTGCGCCGTCCAGCAGCAGC 74874

Qy        694 TCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCG 753
              || ||| ||    |  ||   |||      | || |    ||    ||||  ||| | | |||
Db      74875 TCCGCGCCGTCGGCGTCCTCGTCGTCTGCGCCGTCCAGCAGCAGCTCCGCGCCGTCGGCG 74934

Qy        754 TGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCG 810
              | |||  || |   | || | | |       ||||||||  ||||     ||  | ||
Db      74935 TCCTCGTCGTCTGCGCCGTCCAGCAGCAGCAGCTCCGCGCCGTCGGCGTCTTCGTCG 74991
```

/note="rrnA, 5S ribosomal RNA. Coordinates based on comparison with rrnB from Streptomyces coelicolor (EMBL:AL109848)."
gene complement(5412. .6080)
/gene="SCO4123"
/note="synonym: SCD72A.09c"
CDS complement(5412. .6080)
/gene="SCO4123"
/note="SCD72A.09c, possible two-component system response regulator, len: 222aa; similar to many eg. TR:O86632 (EMBL:AL031155) possible two-component regulator from Streptomyces coelicolor (222 aa) fasta scores; opt: 973, z-score: 1128.5, E(): 0, 69.1% identity in 220 aa overlap. Contains Pfam match to entry PF00072 response_reg, Response regulator receiver domain, Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family and Prosite match to PS00622 Bacterial regulatory proteins, luxR family signature."
/codon_start=1
/transl_table=11
/product="putative two-component system response regulator"
/protein_id="CAB92376.1"
/db_xref="GI:8052435"
/db_xref="GOA:Q9KYE7"
/db_xref="InterPro:IPR000792"
/db_xref="InterPro:IPR001789"
/db_xref="UniProt/TrEMBL:Q9KYE7"
/translation="MTTVLIADDQPLQRFGFRMLLESQDDMTVVGEAANGAEAARMAAELHPDVVLMDVRMPGLDGIEATRRIVAGGGRSRVLILTTFDLDEYAYSGLRAGASGFLVKDALPEELLSGIRAVATGDAVVAPSLTRRLLDAYVHHLPAAGTRDHTPGSPDPRLSPLTERETEILRVMGQGWTNTEIAERLHLAESTVKTHVGRILAKTGSRDRVQAVILAYETKLVG"
misc_feature complement(5415. .5609)
/gene="SCO4123"
/note="Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family, score 61.00, E-value 2.6e-14"
misc_feature complement(5475. .5558)
/gene="SCO4123"
/note="PS00622 Bacterial regulatory proteins, luxR family signature"
misc_feature complement(5736. .6077)
/gene="SCO4123"
/note="Pfam match to entry PF00072 response_reg, Response regulator receiver domain, score 126.30, E-value 5.7e-34"
gene complement(6077. .7447)
/gene="SCO4124"
/note="synonym: SCD72A.10c"
CDS complement(6077. .7447)
/gene="SCO4124"
/note="SCD72A.10c, possible two-component sensor kinase, len: 456aa; similar to many eg. TR:O86631 (EMBL:AL031155) putative two-component sensor kinase from Streptomyces coelicolor (429 aa) fasta scores; opt: 929, z-score: 995.1, E(): 0, 42.7% identity in 422 aa overlap. Contains Pfam match to entry PF00512 signal, Histidine kinase and possible membrane-spanning hydrophobic regions. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative two-component sensor kinase"
/protein_id="CAB92377.1"
/db_xref="GI:8052436"
/db_xref="GOA:Q9KYE6"
/db_xref="InterPro:IPR003594"
/db_xref="InterPro:IPR011712"
/db_xref="UniProt/TrEMBL:Q9KYE6"
/translation="MPATVIPGYYAPHLIYGPRTDARPARTPDHGRMEPVKPVPEDRAAEPVTEYVQRVNRRIRAFDHRHPLLWDLTLTGFWVLAALVDGTGDGWRSVAHDPGVPVGLVLVLSLALSVPVVWRRSHPRGALLLTLPAALVSAWSGAALQAALLQLLLVHHIALRLPLRNLWWAVAVVAAPTCVSVARYGDSAGSWNQQLGSQLVAIGVAALIGITVRTRRGYTEALEDRARRLETERDQQVRLAAAAERARIAREMHDIIGHNLSVITGLADGGRYAAAKSPDRAAQALDAIGTTSRQALGELRRLLDVLRDDDQDGTGPEPELAPQPALTDLDRLLDGVRRAGLPVRTTVKGEPALPEGRQLTLYRVIQEALTNTLKHAGPDATAHIELSYGSGGAATLTVTDTGRTAPLPSRTPSGGRGLPGMRERTSLYGGTLEAGPRPAPEQGWRVHLHLPEESPQ"
misc_feature complement(6095. .6670)
/gene="SCO4124"
/note="Pfam match to entry PF00512 signal, Histidine kinase, score -41.40, E-value 0.0087"
gene 7473. .8006
/gene="SCO4125"
/note="synonym: SCD72A.11"
CDS 7473. .8006
/gene="SCO4125"
/note="SCD72A.11, possible acetyltransferase, len: 177 aa; weakly similar to many eg. TR:AAD52985 (EMBL:AF140221) aminoglycoside 6'-N-acetyltransferase from Xanthomonas maltophilia (153 aa) fasta scores; opt: 128, z-score: 164.9, E(): 0.1, 28.7% identity in 115 aa overlap. Contains Pfam match to entry PF00583 Acetyltransf, Acetyltransferase (GNAT) family."
/codon_start=1
/transl_table=11
/product="putative acetyltransferase"
/protein_id="CAB92378.1"
/db_xref="GI:8052437"
/db_xref="GOA:Q9KYE5"
/db_xref="InterPro:IPR000182"
/db_xref="UniProt/TrEMBL:Q9KYE5"
/translation="MIRSIHADEWPAVKELRLRALRDPVADLAYLETYEEAAARPDSYWQERAARGAEGAAGAQQVVAVGPGGLWLGTLTVLVEEAGTTDWAGFPVERRQGHVVGVFVRPEWRGSGLTKALFDAGLEWGWANGVERMRLIVHPGNTRAQGAYRKAGFAPSGRTVPLEGGPGDHELEYVLER"
misc_feature 7476. .7934
/gene="SCO4125"
/note="Pfam match to entry PF00583 Acetyltransf, Acetyltransferase (GNAT) family, score 34.50, E-value 2.4e-06"
gene complement(8054. .9541)
/gene="SCO4126"
/note="synonym: SCD72A.12c"
CDS complement(8054. .9541)
/gene="SCO4126"
/note="SCD72A.12c, possible membrane protein, len: 495aa; similar in regions to others from actinomycetes eg. TR:Q9X8A2 (EMBL:AL049645) hypothetical protein from Streptomyces coelicolor (427 aa) fasta scores; opt: 266, z-score: 260.7, E(): 4.7e-07, 27.7% identity in 296 aa overlap. Contains possible membrane-spanning hydrophobic region."
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB92379.1"
/db_xref="GI:8052438"
/db_xref="UniProt/TrEMBL:Q9KYE4"
/translation="MRPDDRRDQRPAQGGIPDGLLIGILAFLLGMTLLVWTATGLSALFAHGAWPDGVSFTGTPQAMRQLIAEPHDIPGAWPDAPEGALSGYGLFWGLFIGQLMVLVVLTVFVMGTVARWRAVRQRHREQPTPAPAPAPPAAPVAQETPVQEPPRHEVPTPRAPHQDPALETEPPASAANGKRLGGWETHRAEGAVHYGPGESRHSTAVQTLRDAEGAALVITSNPALWSETKDARAKLGPTHLYDPNHLCDTPARVHWSPTGGCEDRQTATARAAALLTPVRPTARLDQAIGDTAELLLRSYLHAAAIDGRTIRHVHRWSQGLQVQDAVRILRTHPKAAPGSAGELEGSLTAHPERRDMAQQLTARALAALSTVNIREACTPNRTDALALDSFVHEQGTLYVVGESIEDPRTSPGAMPLLTALVSSVVERGRRMAERSSSGRLDPPMTLVLDDVAAVAPLPQLPDLLAGGTDAGMPTLALLRSREQGRARWPHDELPV"
gene complement(9615. .11030)
/gene="SCO4127"
/note="synonym: SCD72A.13c"
CDS complement(9615. .11030)

Query Match 11.3%; Score 101.6; DB 1; Length 299050;
Best Local Similarity 45.3%; Pred. No. 2.9;
Matches 408; Conservative 0; Mismatches 489; Indels 3; Gaps 1;

```
                     LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                     LSPDGLCKTFDADANGYVPGDGVGVLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                     ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                     RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                     VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                     RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                     DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                     RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                     RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                     ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                     RAQFTFKKFLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                     LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                     PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                     PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                     EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE

  Query Match             100.0%;  Score 901;  DB 1;  Length 84868;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-58;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82574 ACGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGC 82633

Qy         61 TCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82634 TCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGC 82693

Qy        121 ACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82694 ACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACC 82753

Qy        181 CGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82754 CGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCC 82813

Qy        241 ACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82814 ACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCC 82873

Qy        301 AGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82874 AGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCC 82933

Qy        361 TCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82934 TCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGG 82993

Qy        421 CGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82994 CGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGG 83053

Qy        481 ACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83054 ACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGC 83113

Qy        541 CGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83114 CGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGA 83173

Qy        601 TCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83174 TCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCC 83233

Qy        661 TCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83234 TCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATC 83293

Qy        721 CCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83294 CCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCG 83353

Qy        781 CCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83354 CCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCG 83413

Qy        841 GCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83414 GCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCGG 83473

Qy        901 A 901
              |
Db      83474 A 83474
```

```
RESULT 2
SCO939119
LOCUS       SCO939119             299050 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 16/29.
ACCESSION   AL939119 AL160331 AL353816 AL356612 AL357432 AL357524 AL358692
            AL358818 AL359988 AL391406 AL391541 AL392149 AL392150 AL392178
            AL645882
VERSION     AL939119.1  GI:24427855
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
   PUBMED   12000953
REFERENCE   2  (bases 1 to 299050)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 30, 2002 this sequence version replaced
            gi:7242748, gi:7636006, gi:8248766, gi:8249971, gi:8388712,
            gi:8439477, gi:8894718, gi:9857143, gi:9886716, gi:10129746,
            gi:11595528.
FEATURES             Location/Qualifiers
     source          1. .299050
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     gene            200. .1727
                     /gene="rrnA"
     rRNA            200. .1727
                     /gene="rrnA"
                     /note="rrnA, 16S ribosomal RNA. Coordinates based on
                     comparison with rrnB from Streptomyces coelicolor
                     (EMBL:AL109848)."
     gene            2004. .5126
                     /gene="rrnA"
     rRNA            2004. .5126
                     /gene="rrnA"
                     /note="rrnA, 23S ribosomal RNA. Coordinates based on
                     comparison with rrnB from Streptomyces coelicolor
                     (EMBL:AL109848)."
     gene            5228. .5348
                     /gene="rrnA"
     rRNA            5228. .5348
                     /gene="rrnA"
```

GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
CDS complement(1992. .3440)
/note="ORFu2; similar to Pseudomonas putida KT2440 PP4696; putative nitrogen regulation protein NR(I)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="AAY32957.1"
/db_xref="GI:63099934"
/translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
ELTETEIARRLGISRKALWERRQRLGIPRPRS"
CDS 3587. .5548
/note="ORFu3; ORFu3p; similar to Leptospira interrogans AAN50474"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAY32958.1"
/db_xref="GI:63099935"
/translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
VVAGFGATGPLRVMETLPEQYLFLTRVRP"
CDS complement(5702. .7231)
/note="ORFu4; ORFu4p; similar to Burkholderia fungorum hypothetical protein in ZP_00033665"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAY32959.1"
/db_xref="GI:63099936"
/translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPFSANQRPASTLAQWRVGLEHP
GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
GRVITVHFHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
P"
CDS complement(7561. .8931)
/note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440 PP4696; putative nitrogen regulation protein NR(I)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="AAY32960.1"
/db_xref="GI:63099937"
/translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
PPPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
IPRPRS"
CDS complement(8928. .11939)
/note="ORF1; ORF1p; similar to Pseudomonas putida KT2440 HisK (PP4695) in GenBank Accession Number AAN70268"
/codon_start=1
/transl_table=11
/product="putative sensory box histidine kinase"
/protein_id="AAY32961.1"
/db_xref="GI:63099938"
/translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
FTAFFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPAREEREAAR
LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
CDS complement(12405. .13379)
/note="ORF2; ORF2p; similar to family of known or putative phosphotransferases, including macrolide 2-phosphotransferases; similar to Bacillus subtilis YcbJ, Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273 MphB and Staphylococcus aureus plasmid pSR1 MphBM"
/codon_start=1
/transl_table=11
/product="putative macrolide kinase"
/protein_id="AAY32962.1"
/db_xref="GI:63099939"
/translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
CDS 13368. .14663
/note="ORF3; ORF3p"
/codon_start=1
/transl_table=11
/product="putative Ser/Thr protein kinase"
/protein_id="AAY32963.1"
/db_xref="GI:63099940"
/translation="MFGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
gene 15731. .33766
/gene="dszA"
CDS 15731. .33766
/gene="dszA"
/note="PKS (polyketide synthase) subunit modules 1-4a"
/codon_start=1
/transl_table=11
/product="DszA"
/protein_id="AAY32964.1"
/db_xref="GI:63099941"
/translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
PRAHYSPDFNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 4992.51 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_75000_75900
Perfect score: 901
Sequence: 1 acggccgggacgcggacgac.....gaggaggatcgcgccgcgga 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

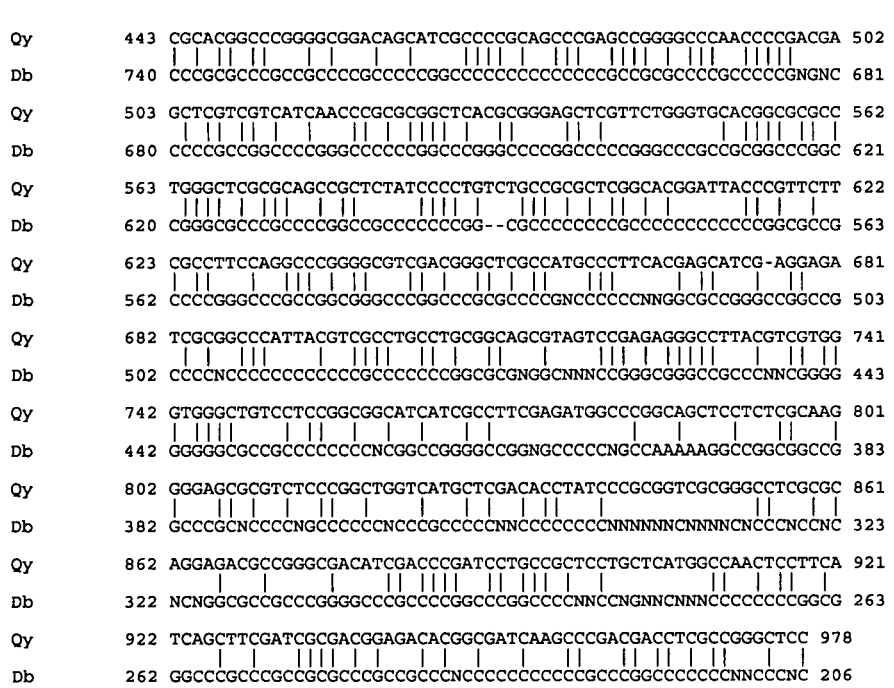
SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	901	100.0	84868	1	DQ013294	DQ013294 Polyangiu
2	101.6	11.3	299050	1	SCO939119	AL939119 Streptomy
3	98.6	10.9	110000	2	CP000081_01	Continuation (2 of
c 4	97.8	10.9	293050	1	SCO939116	AL939116 Streptomy
c 5	97	10.8	22437	1	AY027524	AY027524 Frankia s
c 6	95.6	10.6	48000	1	POL505006	AJ505006 Sorangium
7	95.4	10.6	1926	6	AR217867	AR217867 Sequence
8	95.4	10.6	1931	6	AR083152	AR083152 Sequence
9	94.8	10.5	300100	1	SCO939123	AL939123 Streptomy
10	91.4	10.1	1869	6	AX470092	AX470092 Sequence
11	91.4	10.1	2126	6	AX470093	AX470093 Sequence
12	91.4	10.1	3780	6	AX470091	AX470091 Sequence
c 13	91.4	10.1	110000	1	BA000030_47	Continuation (48 o
c 14	91.4	10.1	299050	1	SCO939119	AL939119 Streptomy
c 15	91	10.1	321250	1	SCO939111	AL939111 Streptomy
c 16	90.8	10.1	110000	1	BA000030_23	Continuation (24 o
17	90.6	10.1	295150	1	SCO939126	AL939126 Streptomy
18	90.4	10.0	110000	2	CP000081_02	Continuation (3 of
19	89.6	9.9	110000	15	AP008217_032	Continuation (33 o
c 20	89.6	9.9	154441	15	AC120527	AC120527 Oryza sat
c 21	89.6	9.9	161073	15	AC147812	AC147812 Oryza sat
c 22	89.2	9.9	15898	6	BD262986	BD262986 DNA seque
c 23	89.2	9.9	15898	6	AX024286	AX024286 Sequence
24	88.8	9.9	32748	1	AB070951	AB070951 Streptomy
c 25	88.8	9.9	110000	1	BA000030_45	Continuation (46 o
c 26	88.4	9.8	77534	1	AF235504	AF235504 Streptomy
c 27	88.4	9.8	77536	6	BD235937	BD235937 Polyketid
c 28	88.4	9.8	77536	6	AR271638	AR271638 Sequence
c 29	88.4	9.8	77536	6	AR564393	AR564393 Sequence
30	87.8	9.7	22533	6	CQ875929	CQ875929 Sequence
31	87.8	9.7	110000	14	LMFLCHR32_06	Continuation (7 of
32	87.8	9.7	161719	1	AY899214	AY899214 Streptomy
33	87.8	9.7	164051	6	CQ875911	CQ875911 Sequence
c 34	87.2	9.7	9589	1	SCO001205	AJ001205 Streptomy
35	87.2	9.7	90600	1	AJ862840	AJ862840 Streptomy
c 36	87	9.7	82746	1	AF453501	AF453501 Actinosyn
c 37	87	9.7	310550	1	SCO939113	AL939113 Streptomy
38	86.8	9.6	27541	6	AX211706	AX211706 Sequence
39	86.8	9.6	123580	1	AF263912	AF263912 Streptomy
40	86.8	9.6	125401	6	AX211739	AX211739 Sequence
c 41	86.6	9.6	110000	14	CT005244_0	CT005244 Leishmani
42	86.4	9.6	22449	1	SPSNBCDE	Y11548 S.pristinae
43	86.4	9.6	22449	1	SPSNBCGEN	X98690 S.pristinae
c 44	86.2	9.6	110000	15	AP008209_131	Continuation (132
45	86.2	9.6	151173	15	AC135257	AC135257 Genomic s

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRAREAEREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPFFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

THIS PAGE BLANK (USPTO)



Search completed: April 28, 2006, 12:17:49
Job time : 4958.14 secs

/clone_lib="NIH_MGC_18"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 11.2%; Score 111.8; DB 5; Length 1046;
Best Local Similarity 48.4%; Pred. No. 8e-10;
Matches 347; Conservative 0; Mismatches 366; Indels 4; Gaps 2;

```
Qy         85 CCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGC 144
Db        985 CCGCCCCGCCNCCCCCCNCCCCGGCGCGCGCGCGGCCGCGCCCGGGGGCGNCGCNNCCGC 926

Qy        145 TCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCC 204
Db        925 NCGCCGCCCCGCCGCCCCGCCGCCCGCGGCGCGCGCCCCGCCCCGCGCGGCCGCCGCCGCC 866

Qy        205 TGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGT 264
Db        865 GGCGCCGCCGCCGNNCCGCGGGCCGCCCGCGGNCCCCCCCGGCCCCCCCCGCCCGGCCGC 806

Qy        265 CGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCT 324
Db        805 CCGCGCCCCCCCCCCCGCCGCGGCCCGCGCCCCCGCGCGGCGCNCCCCCGCCGCCGCGCC 746

Qy        325 CCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGG 384
Db        745 CCGCGCGCCCCGCCGCCGCCNCGCGCCCCCGGCCGCGCGCCGCCCCGCGCCCCCCCCCCC 686

Qy        385 CCTCCGCAGTCCCGGCG---GCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGG 441
Db        685 CCCCCGCGCCCGCCGCGCGCCCCCCCCCCCGCGCGGGCGCCGCGGCCCGGGGCGCCGCCC 626

Qy        442 ACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACG 501
Db        625 CCGCCGCCCCCCCGCCCCCCCCCCGCGGCCCCCCCCCCCGCCCCGCCCCCNCCCCCCCCC 566

Qy        502 AGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGC 561
Db        565 CGCCCGCCGCGGGCGCCGCCCGCCCCCCCCCCCGCGCCGCCCCCNCGCCCGCCCGCGGCC 506

Qy        562 CTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCT 621
Db        505 CCGCGC-CGCGGCGCCGCCCCCCGCCCCCCCCGCGCGCGCCCCCCCGCCCCCCCCCCC 447

Qy        622 TCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGA 681
Db        446 CCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGCCCCCCCCCCCGGCCGCCCGGGCC 387

Qy        682 TCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGG 741
Db        386 CCCGGCCCCGCCCCCCCGCCCGCGGCCCGGGGGGCGGCGCGCGGGCCCCCCCGGCCCGCC 327

Qy        742 GTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGC 798
Db        326 CCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCGCNCGGGGCGCCGGCCGCGCCGCCCCC 270
```

RESULT 15
AG441574/c
LOCUS AG441574 1216 bp DNA linear GSS 22-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-323A15.TJ, genomic survey sequence.
ACCESSION AG441574
VERSION AG441574.1 GI:48084637
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1216)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1216
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-323A15.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 11.2%; Score 111.6; DB 10; Length 1216;
Best Local Similarity 45.4%; Pred. No. 8.6e-10;
Matches 407; Conservative 0; Mismatches 485; Indels 5; Gaps 3;

```
Qy         85 CCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGC 144
Db       1100 CCCGGCCCCGCCGCGCGGCGNGCCCGCCCGGGCCGCGGCCGCCCGCCCCCCCCCCCCCGC 1041

Qy        145 TCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCC 204
Db       1040 CGGCCGGCCCGCCCCCCNCGCCCCGCCCCCCGCCGGCCCCCGCGCCCGGCCCGCCGCCCC 981

Qy        205 TGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGT 264
Db        980 GGCCCCCCCCCCGCCCGCCCCGGCGCGCCCCCGCCCCCGCCCCGNGCGCGCGCCCCCCCC 921

Qy        265 CGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCT 324
Db        920 GGCCCCCCGCCGGCGCCGCCCGCCCCGCGGCCGCGCCGCCGGGCCCCGCCCCCCCGGCCG 861

Qy        325 CCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGG 384
Db        860 GCCCCCCCGGGGCGCCCCGCCGCCGGCCCCGGCCGGCCCCCCCCCCCGGGGGCGCCCCCC 801

Qy        385 CCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGG--GGA 442
Db        800 CCGCCCCCCGGCCGGGCCCCGCCCCCGGCGCCCCGCGNGCGGGCCCCCCCCCCGGCCGCC 741
```

```
Db        366 CGCNCGCCCGCNGCCNNCCCGCCGCCCCCCGCCNGGGCCCCCGGCGGCGCCCNGCCNNCC 307

Qy        750 TCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGC 809

Db        306 GCGGCCCGCCCCCCGCGCGCCCCCCGCGCGGCCCNNGGGGCCCNGC------CGCCCGCCC 253

Qy        810 GTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACG 869

Db        252 GCCNGCCGGCCGGCGCCCCGCGCCGCCCCCCCCCCCCCGGGNGGCGGCCCCCCCCGNGCC 193

Qy        870 CCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTC 929

Db        192 CGNNGCCCCGGGNNNNCNCNCNNCNNCNNCCGCCCGCNNCCCCNCNCNNNNNNCNNCNNN 133

Qy        930 GATCGCGACGGAGACAC 946

Db        132 NCNNNNNNNCGGGAAAC 116


RESULT 13
BM804596
LOCUS       BM804596               1153 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6492847 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554728
            5', mRNA sequence.
ACCESSION   BM804596
VERSION     BM804596.1  GI:19121419
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1153)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12273  row: m  column: 01
            High quality sequence start: 53
            High quality sequence stop: 352.
FEATURES             Location/Qualifiers
     source          1. .1153
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5554728"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_85"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             11.2%;  Score 112;  DB 3;  Length 1153;
  Best Local Similarity   50.5%;  Pred. No. 7.4e-10;
  Matches  284;  Conservative    0;  Mismatches  275;  Indels    3;  Gaps    1;

Qy         88 GCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCA 147

Db        465 GCCCCCCCGCCCCCCGCCGCCCCCCGCCCCCGCCCCGGCCCCCCCCGGCCCCGCCCCGGC 524

Qy        148 GCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGA 207

Db        525 CCCCGGCGCCCCGCGGGGCCCCCCCCGCCCCCGCCCCCGCGCCCCGCCGGCGCGCGCCCG 584

Qy        208 GCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGA 267

Db        585 GGCCCCGCCCGCCCGCCCCCCGCCCCCCCGCCCGGCCNNCCGGCGGCCCCCGGCCCGCGC 644

Qy        268 TCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCG 327

Db        645 CGCCCGCCGCCCCCGCGGCCCCCCCCCCCCCCCGCCCGCGGGCCGGCCGCCCGGCCGGCC 704

Qy        328 CAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCT 387

Db        705 CGCGCCCCCGCCCCGCCCCCCGGGGGGCCGCCGCGGCGGGCCGCCCCCCCGCCCGGGGNG 764

Qy        388 CCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCAC 447

Db        765 CGCCCGCCGGGCGCCCCCCCCGCGCCGCGCGGGCCCCCCGCGCNCGGGCGCGGGCGGCCC 824

Qy        448 GGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCG 507

Db        825 GGGCCGCGGCCCCCGGGCCCGCGCCCCCGCCCGCCCCGCGGCGCGCCCGCCCCCGGCNCG 884

Qy        508 TCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGC 567

Db        885 CCGCCCGCCGGCCGCGCCCCCCGCCCCCCCGCCCGCCCGCGGCG---GGCCCGCCCGCCG 941

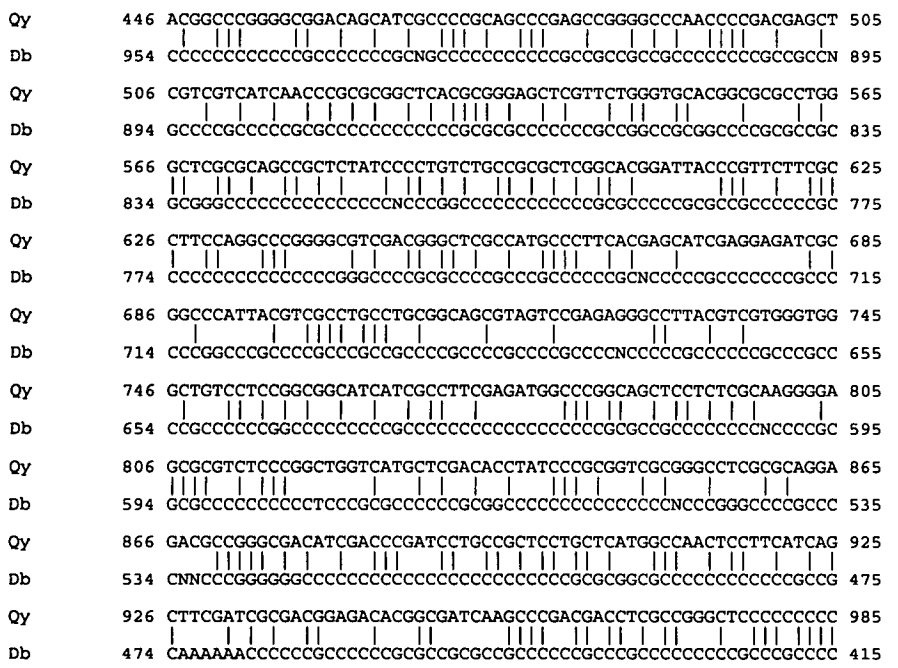
Qy        568 TCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCT 627

Db        942 CCCCCCCGCGCCCCCGCCCCCGCGCCNCCGGCGCCGGCGGTTCCNNCNCCCNCGCCGCCG 1001

Qy        628 TCCAGGCCCGGGGCGTCGACGG 649

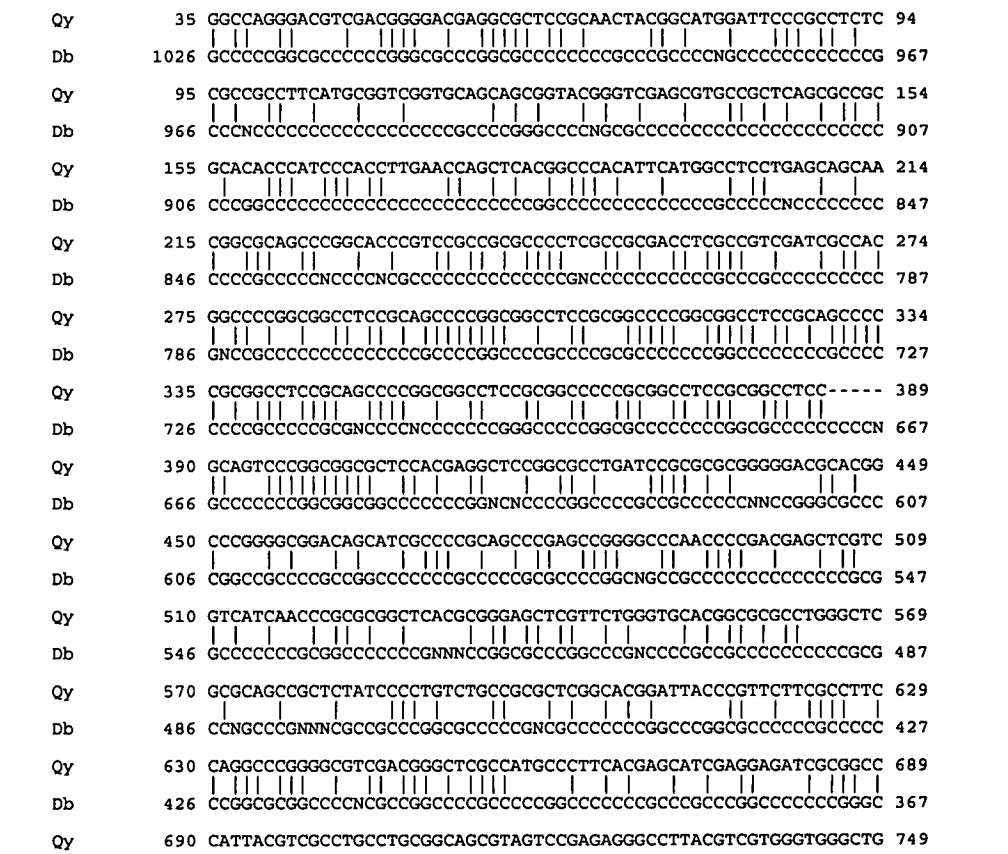
Db       1002 GTGTCGCCCGCGCCGCCGCCGG 1023

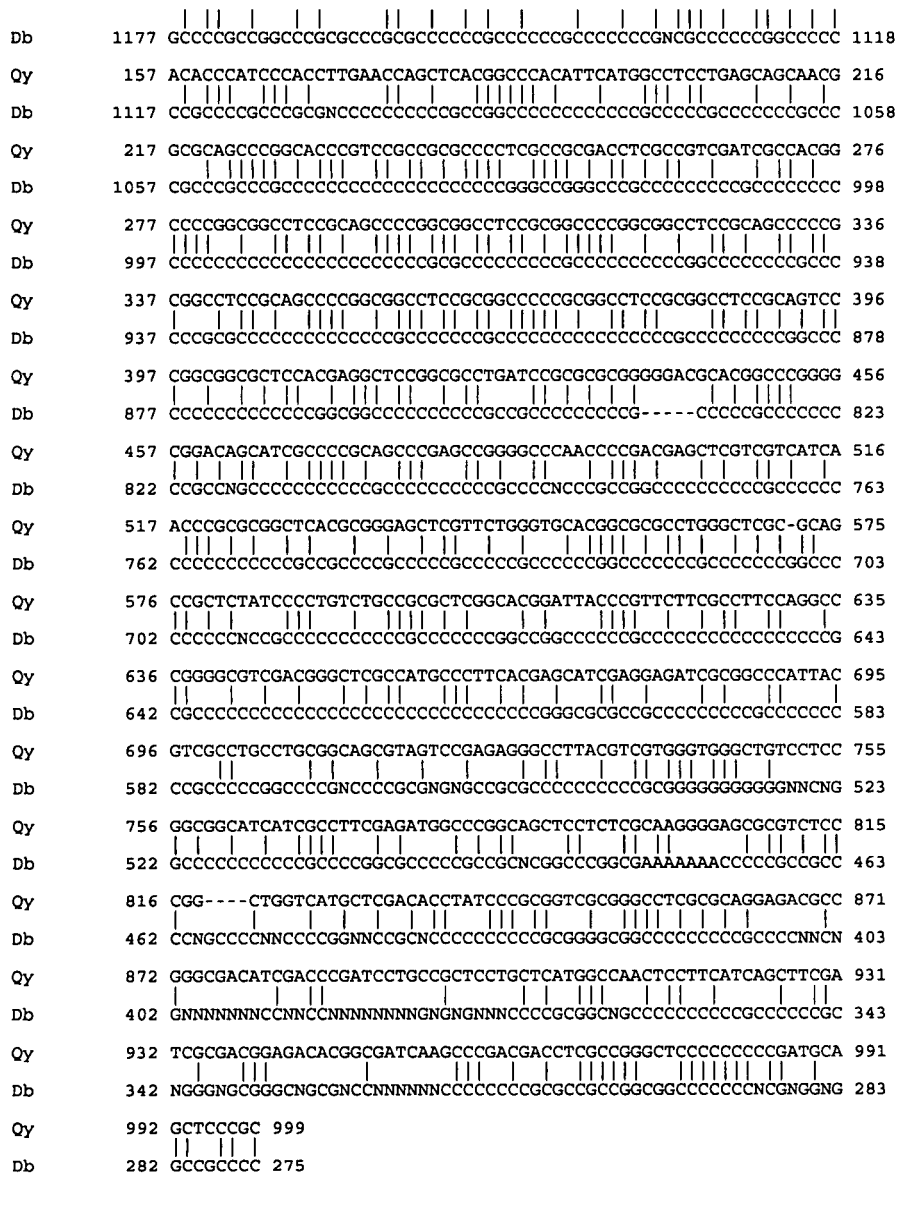

RESULT 14
BQ952554/c
LOCUS       BQ952554               1046 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8863775 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423858
            5', mRNA sequence.
ACCESSION   BQ952554
VERSION     BQ952554.1  GI:22368032
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1046)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2603  row: b  column: 19
            High quality sequence stop: 206.
FEATURES             Location/Qualifiers
     source          1. .1046
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6423858"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
```



```
RESULT 12
AG448422/c
LOCUS       AG448422               1286 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-332F23.T7, genomic survey
            sequence.
ACCESSION   AG448422
VERSION     AG448422.1  GI:48091485
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1286)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1286
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-332F23.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

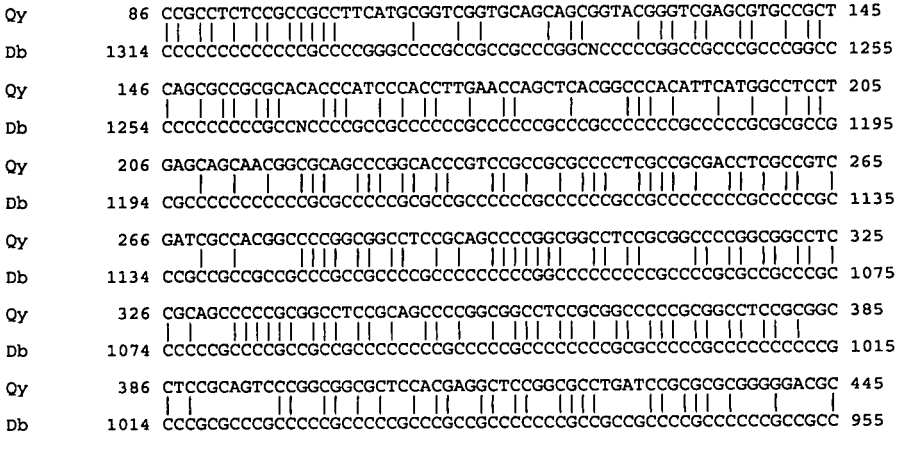
Query Match 11.3%; Score 112.6; DB 10; Length 1286;
Best Local Similarity 43.7%; Pred. No. 5.7e-10;
Matches 401; Conservative 0; Mismatches 505; Indels 11; Gaps 2;





RESULT 11
AG448267/c
LOCUS AG448267 1523 bp DNA linear GSS 22-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-332C15.TJ, genomic survey sequence.
ACCESSION AG448267
VERSION AG448267.1 GI:48091330
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1523)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1523
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-332C15.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 11.3%; Score 113.2; DB 10; Length 1523;
Best Local Similarity 44.8%; Pred. No. 4.4e-10;
Matches 409; Conservative 0; Mismatches 504; Indels 0; Gaps 0;



```
          High quality sequence start: 26
          High quality sequence stop: 136.
FEATURES             Location/Qualifiers
     source          1. .1695
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-172N5"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             11.7%;  Score 117;  DB 9;  Length 1695;
  Best Local Similarity   44.6%;  Pred. No. 9.5e-11;
  Matches  390;  Conservative    0;  Mismatches  484;  Indels    0;  Gaps    0;

Qy        125 GTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCAC 184
              || || |   |  |  |  |  |     | |||  ||| |       ||  | | |
Db        121 GTCCGAGGGAATGGCTCNNCCNNCCNGCCNCTCCCCCCCCCCCCCCNNCCCCCCGCCCCCC 180

Qy        185 GGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCC 244
                | |    |  | || ||    |  |  |  | |  ||   | |||| ||||| | ||
Db        181 CNCGCGGCCCCCCGCCCGCCCCGCCCCCCCCCCCCCCCCCCGCCCCCCCGCCCGCCCCCC 240

Qy        245 CCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGC 304
              || | || |  || | ||      |    |  ||||  |  || ||||  |||   ||||
Db        241 CCCCCCCCCCCCCCCCCCCCCGCCCCGGCCCCCCCCCCCCCCCCCCCGCCCCCCGCCCGGC 300

Qy        305 CTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGC 364
              | ||| |  || |  || || |  || ||||  || ||||  ||||  |  || || | |
Db        301 CCCCCCCCCGCCCCCCCCCCCCCCCGCCCGCCGCCCCCCCCCGCGCCCCCCCCCCCCCCCC 360

Qy        365 GGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCC 424
                |||||  |  || || |  || || |  ||| ||  |  || |    | ||  | ||
Db        361 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCGCCCCCCCCCCCCCCCCCC 420

Qy        425 TGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCC 484
                || | | |      | | |  |||    |  |  |  | |||| |  |||   ||
Db        421 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 480

Qy        485 GGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTT 544
                |||  |||| |  | | |  |  | | |  ||| | |  | | | |    | |
Db        481 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCGC 540

Qy        545 CTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGG 604
              |    || |  ||| ||  | | | | |  || | |   ||||   |  || | | |
Db        541 CCCCCCGCCCCCCGCCCCCCGCCGCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCC 600

Qy        605 CACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTT 664
              | |      ||    | | ||  ||  ||||    |  |  ||  | | ||   |||
Db        601 CCCCCCCCCCCGCCCCCCCCCCCCCCCCCGCCCCCCCCNCCCCCGCCCCCCCCCCCCCCC 660

Qy        665 CACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAG 724
              | |  ||  ||  |  | ||  ||    || | ||  |      |  |     | |  |
Db        661 CNCCCGCGCCGCCGCCCCCCCGCGCCCCNGCGCCCCCNCCGCCGCCCGCCCCGCCNCCCG 720

Qy        725 AGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCG 784
                ||||  ||  |    |  |    ||  |  |  |  | ||  |    | |||
Db        721 CCGGCCGCCCGCCCGCCCCGCNCCCCGNCCCCCCCCCGCCGCCCCCGCCCCCCCGCCCCC 780

Qy        785 GCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGC 844
              ||  | || | | |  |     | || | |||  |     |  | ||  |      ||  |
Db        781 GCCCCCCCCCGCCCCGGCCCCCCCCGGCCCCCNCCCCNGCCCCCCCGCCCNGCCCCCGNC 840

Qy        845 GGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCT 904
                | | |||| | | |  | | | |  |  | |  | |||   ||  || | ||  |
Db        841 CCCCNCNGGCCCCCCCCCGCGCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 900

Qy        905 CATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCGACGA 964
              |    ||  | ||      |  |  |      || |      | |   |  |  |||  |
Db        901 CCCCCCCCCCCCCCCGNCGCNNCCCCCCCNCCGCCCCCCCCCCCCCNGCCCCCCCCCNNCCC 960

Qy        965 CCTCGCCGGGCTCCCCCCCCCGATGCAGCTCCCG 998
              || ||||   | |||||||||   ||  | ||||
Db        961 CCCCGCCCCCCCCCCCCCCCCCCCGCCCCCCCCG 994


RESULT 10
AG441757/c
LOCUS       AG441757               1425 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-323E17.TJ, genomic survey
            sequence.
ACCESSION   AG441757
VERSION     AG441757.1  GI:48084820
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1425)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1425
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-323E17.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             11.5%;  Score 115.4;  DB 10;  Length 1425;
  Best Local Similarity   44.9%;  Pred. No. 1.8e-10;
  Matches  435;  Conservative    0;  Mismatches  523;  Indels   10;  Gaps    3;

Qy         37 CCAGGGACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCG 96
              || ||  |  |  |  |  |   || | || |  |  | ||       ||| || | ||
Db       1237 CCCGGCGCCCCCGCCCGCCCCCCGCCCCCCCCCCCGCCCGCCCCCCCCCCCCCCGCCCCC 1178

Qy         97 CCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGC 156
```

```
RESULT 8
AG396765/c
LOCUS       AG396765               1692 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-217C03.TJ, genomic survey
            sequence.
ACCESSION   AG396765
VERSION     AG396765.1  GI:48030934
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1692)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1692
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-217C03.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             11.7%;  Score 117;  DB 10;  Length 1692;
  Best Local Similarity   47.0%;  Pred. No. 9.5e-11;
  Matches  409;  Conservative    0;  Mismatches  456;  Indels    5;  Gaps    2;

Qy        131 GTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCA 190
              | |  ||  ||||| |  |||||| | | |||   |  |   |  || || | ||  ||
Db       1502 GCCCCGCCCGCCGCGCCCCGCCGCNCGCCCCCGCGCGCCGCCGCCCCCGCCCCCGCGCCC 1443

Qy        191 CATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGC 250
                  |    || |  | ||  |      ||| |||| || | |  || |||| ||||   |
Db       1442 GCGCCCCCCCCGCGCGCGCCCCGCGCCCGCCGCCCCGCGCGCCGCCCCCGCCCCCCCGCC 1383

Qy        251 CGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGC 310
              | |  || | ||   |  | || |  || || |  || ||   |||||  | ||  |  |
Db       1382 CCCCCCCCCCCCCCGGCCCCCCCCCCCCGCGCCCCCCNCCCGCGCCCCCCCCGCGCCGNC 1323

Qy        311 GGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 370
```

```
              |||   |  || ||||  |||||||  ||  | | |||||  |  || |     |||
Db       1322 CCCCCGCCCCCCCCCCCGCCCCCCCCGCCCCCCGCCCCGCCCCCCCCCCCCCGCGCGCCCCCG 1263

Qy        371 CGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCC 430
              | |  || | ||  || || |  |||  |  ||| || |   || ||  | ||    ||
Db       1262 CCCCCCCCCGGCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCGCCCCCCCCCCCCCCCC 1203

Qy        431 GCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCC 490
              || | ||     |   |  ||| | ||   | ||    |||| |  |||  |||    ||
Db       1202 GCCCCCGCCCCCCCGCCCCCCCCCGCGCCCCCGCCCGCCCCCCCCCCCCCCGCCCCCCC 1143

Qy        491 CAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGT 550
              |    ||   ||  | |  || |  |  ||||| || | | |||    || |   |
Db       1142 CGCGCCNCGCGCCCCCCCCGCCCCCCCCCCGCCCGCCGCCCGCCCCCGCCC---CCCCCC 1086

Qy        551 GCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGA 610
              || || |||||    || | | |  |  | |  | ||   | |||  | | | || |
Db       1085 GCCCGCCGCGCGCCCGCCCCCCCGCCGCCCCCCGCGCCCCCCCGCCCCCCCCCGCGCCCC 1026

Qy        611 TTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAG 670
              |||   |   ||  |  ||||  ||  |  |   | ||   |||  | |
Db       1025 GCGCCCCCCCCCGCCCCGCNCNCCCCGCCGCCGCCGCCCCGCCCCCCCCCCCCCCCCCC 966

Qy        671 CATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCC 730
              ||        | ||||||    || | ||  || ||| | ||     |||    ||||
Db       965 GCGCGCCCCCCCCCCGGCCCCCCCCGCCCCC--CCCGCGCCGGCCCCCCCCGCCCCGGCC 908

Qy        731 TTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCT 790
              || |       | | | || ||| | ||  |  |||| | |  ||||  | || |  |
Db       907 CCCCGCCCGCCCGCGCCCGCCCGCCGCCCGCCGCCNCGCCCCCCCCGCGCGCCCCCCCGCC 848

Qy        791 CCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGC 850
              ||  | |  |    | |  | | |  |       |  |  |      | |||   | |
Db       847 CCCNCCCCCCGCCNCCCCCCGCCCGCCCCCCCCCGCNCCCCCCCCGGCCCCGCGCCCCCCC 788

Qy        851 GGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGC 910
                | || | ||       | || |||| |  |  |||   ||  || | |  || |  |
Db       787 CGCCCCCCCGGCCCCCCCCCCCGGCGCCCCCCCCCCCCCCCCCCCCCCGCGCCCCCGCG 728

Qy        911 CAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCGACGACCTCGC 970
              |    ||  |  |    ||  | |  |      |   ||   |  |||| |  ||   |
Db       727 CCCGCCCGGCCCCGCGCCCGCCCCCCCCCCGCCCGCGCCCGCCCCCCCGCCCCCCCNCC 668

Qy        971 CGGGCTCCCCCCCCCGATGCAGCTCCCGCG 1000
              |     ||||| |||       | ||| ||
Db       667 CCCCGCCCCCCGCCCCCCCGGCCCCCCCCG 638
```

```
RESULT 9
CC290874
LOCUS       CC290874               1695 bp    DNA     linear   GSS 13-MAY-2003
DEFINITION  CH261-172N5_Sp6.1 CH261 Gallus gallus genomic clone CH261-172N5,
            genomic survey sequence.
ACCESSION   CC290874
VERSION     CC290874.1  GI:30662315
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1695)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
```

```
Db       1143 CCCGCCGCCCCCGCCCGCCCCCGCGCCCCCGCCCGCCGCCGCGCCCGCCCCNCCCGCGCCC 1202

Qy        681 ATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTG 740

Db       1203 CCCGCGCCCGCCGCGCCCGGCGCCCCCCCCCGCCGCGCGCCCGGCCCCCCCGCCCCGCCG 1262

Qy        741 GGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAA 800

Db       1263 CCCGCGCGCCCCCCCCCGCGCCCGCCCCGCCCCGCCGCCCCCGCCCCCGCCCCCCCGCCGC 1322

Qy        801 GGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCG 860

Db       1323 CGCCCCCCGGCCCCCCGCCCCGCCCGCCCGCGCCCCCCGGCCCCCNCCCCCCCGCCCGCG 1382

Qy        861 CAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTC 920

Db       1383 NCGCCGCCCCGCGCCGCCCGCCGCCCGGCCCCGCCCCCGCCGCCGCCCCCCGCCCCCCNC 1442

Qy        921 ATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCGACGACCTCGCCGGGCTCCCC 980

Db       1443 GGCGCCGCCCGCCCCCGCGCCCGCGCCCCCGCCCGGCCCCGCGCGCCGGCCGCGCCGCCC 1502

Qy        981 CCCCCGATGCAGCTC 995

Db       1503 GCGCCGCCCCCGCGC 1517


RESULT 7
AG435283/c
LOCUS       AG435283               2017 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-313F03.TJ, genomic survey
            sequence.
ACCESSION   AG435283
VERSION     AG435283.1  GI:48078346
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 2017)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .2017
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-313F03.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match           11.8%;  Score 118.4;  DB 10;  Length 2017;
  Best Local Similarity  47.8%;  Pred. No. 5.3e-11;
  Matches  404;  Conservative    0;  Mismatches  429;  Indels   13;  Gaps    2;

Qy        151 CCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCA 210

Db       2017 CCGCGCCCCCCCCCGCCCGCGCCGCGCCCCCGGCCGCCGCCGCCCGCGCCGCGCCCCCCC 1958

Qy        211 GCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCG 270

Db       1957 GCCCCGCGCGNGCNCGCCGCGCCGCGGCCGCGCGGCGCCCCCCCGCCCCCGCCCCGCCCG 1898

Qy        271 CCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAG 330

Db       1897 CCCCCCCCCCGCCGCCGCCCCGCGCCCCGGCCCCGCCCCCCCCCCCCGCCCCCGCCCCCC 1838

Qy        331 CCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCG 390

Db       1837 NCCCCGCCGCCGCCCGCCCCCCGCCGCCCGCCGCGCGCCGCCCGCGCCCGCCCCCCCCGCCC 1778

Qy        391 CAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGC 450

Db       1777 CCGCGCGCCCCGCCCGCCCCCGCCCGCCCGCCGCGCCCCGCCGCGCCCCCCCGCGNCGCCGC 1718

Qy        451 CCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCG 510

Db       1717 CCCGCGC----------CCCCGCCGGCCCCGCCGCGCGGCCGCCNCCCCGCCCCCCCCCG 1669

Qy        511 TCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCG 570

Db       1668 CCCGCGCCCCGCCCGCGCCCGCCCCGCCGCCGCCCCCCGCGCCCCGCGCCCCCCCGCCCG 1609

Qy        571 CGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCC 630

Db       1608 GCCCCCCGCCCCCGCCCCGGGCCCGCCCCCGCGGCCCGCCCCCCCGCCCGCGCGCCGCCC 1549

Qy        631 AGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCC 690

Db       1548 GCGCCCCGGCGCCCGCCCCCCCCGGCGCGCGCCCCCCCCGCCGCCCCCCCCCCCCCCCC 1489

Qy        691 ATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGT 750

Db       1488 GCCCGCCCCCCGCCGCCGCCCCCCCCGGCCCCCCGCCCCCCGCCCCCCGCCCCCGCCCG 1429

Qy        751 CCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCG 810

Db       1428 CGCCCCCCCCCCCCCGCGCCCCCCCGCGCCCCGCGCCCGCCCCCGCGCGGCCCGCGCCGCG 1369

Qy        811 TCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGC 870

Db       1368 CCGCGCGCCNGGCCGCCCGCGCCGCC--CCCGCCGCCCGCCGCCCCGCGCCCCCGCGCCGG 1311

Qy        871 CGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCG 930

Db       1310 CGCCCCCCCCCCCCCCCGNCGCGCCCCCCCGCCCGCGCCGCGCCCGCCGCCCCCGCCCCG 1251

Qy        931 ATCGCGACGGAGACACGGCGATCAAGCCCGACGACCTCGCCGGGCTCCCCCCCCCGATGC 990

Db       1250 CGCGCGCCGCCCCCCCCGCCCCCCCGCCCCGGCCGCCCCGCNCCCGGCCCCCCCGCGCGCC 1191

Qy        991 AGCTCC 996

Db       1190 CCCCCC 1185
```

```
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES            Location/Qualifiers
     source         1. .996
                    /organism="Pan troglodytes"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9598"
                    /clone="PTB-063O11.R"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match            11.9%;  Score 119;  DB 10;  Length 996;
  Best Local Similarity   48.6%;  Pred. No. 4.4e-11;
  Matches  305;  Conservative    0;  Mismatches  322;  Indels    0;  Gaps    0;

Qy         36 GCCAGGGACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCC 95
Db        269 GCCCGCGCCCGGGGGGGGGGGGCGCGCGCGCCCCGCGGCCGGGGGGGCCGCGCGCGGTCTC 328

Qy         96 GCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCG 155
Db        329 CCGGCCGCCCGCGGGGCGGCCGCGGCGCGGCCGCGCGGGCGCCGCCCGCCCCCCCCCCGC 388

Qy        156 CACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAAC 215
Db        389 CGGCCCCCGCCCCGGGGGGGGCGGCCGGCGGNNCGGCGCGCGGGGGNGCGGGGCCCCCCC 448

Qy        216 GGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACG 275
Db        449 CGGCCCGCCCCCNCGCGCCCCCCCCCCCGCCCCGCCCCGCCCCCGCCCGCGCGNCCCCCG 508

Qy        276 GCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCC 335
Db        509 CGGCGCGCCCCCGCCGCGGGCCCGCCGGCGCCCGCGCGGCCCGCGCCGCGCCCCGCGGCC 568

Qy        336 GCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTC 395
Db        569 CGCGGCCCGCCCGCCCCGCCCGCCCCGGCGGGCGGCGCGGCCACGCGCCGGAGGGGGGGC 628

Qy        396 CCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGG 455
Db        629 CGGCGCGCGCGGCGCGCCGGCGCGGCGGCGGGGCGGCGCGGCGCCCCAGGGCGGGCGGGC 688

Qy        456 GCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATC 515
Db        689 GCCGCCCCGGCGCGCGCGCGGCGCGCGNNNGCGCANNCGGCGCCCCCGCGCGGCGCCCGC 748

Qy        516 AACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAG 575
Db        749 CCCCCGCGGCGGCGGCGCGGCGCGCCGGCGCCGGGCGGCCGCGCGCCCGCGGCCGCGGCG 808

Qy        576 CCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCC 635
Db        809 GCGCCGCCCCGCGCGCGCGGGGCGCGGCGCGGGCGGGGCGCGCGCGCCGCCGGGCGGGCG 868

Qy        636 CGGGGCGTCGACGGGCTCGCCATGCCC 662
Db        869 CGCGNCGCGCGCCGGCCGGCCNGCCGC 895


RESULT 6
CL078342
LOCUS       CL078342               1674 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-149D22_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-149D22, genomic survey sequence.
ACCESSION   CL078342
VERSION     CL078342.1  GI:40534255
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1674)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 300
            High quality sequence stop: 387.
FEATURES            Location/Qualifiers
     source         1. .1674
                    /organism="Xenopus tropicalis"
                    /mol_type="genomic DNA"
                    /strain="Nigerian frog"
                    /db_xref="taxon:8364"
                    /clone="CH216-149D22"
                    /sex="male"
                    /cell_line="Stock 248 F7A2, inbred N7"
                    /clone_lib="CH216"
                    /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                    BAC library"
ORIGIN

  Query Match            11.9%;  Score 118.6;  DB 10;  Length 1674;
  Best Local Similarity   46.2%;  Pred. No. 5e-11;
  Matches  423;  Conservative    0;  Mismatches  488;  Indels    4;  Gaps    1;

Qy         85 CCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGC 144
Db        603 CCCCCCCCGCCCCCCCGCCCCCCCCGCCGCCCCCGCCCCCCGCCCCCGCGCCGCCCCCCCC 662

Qy        145 TCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCC 204
Db        663 CCCGCCCCCCCCCCCCGCGCCCCCCGCCCCCGCCGCCCCCCCCCCCCCCCCGCCCGCGCCGC 722

Qy        205 TGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGT 264
Db        723 GCGCGCGCCCCGCCGCCCCCCGCCCCCCCCCCCCCCCGCCCCGCCCCCGGCCCCGCGCCCC 782

Qy        265 CGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGG----CG 320
Db        783 GCCCGCGCGCCCCCCCCCCCGCCCCCGCCGCCCCCCCCGCCCCCCCCGCCCCCCCGCCCCG 842

Qy        321 GCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCC 380
Db        843 GCCCCCCCCCCCGCCGCCCCCCGCGCCGCCGCCCGCCCCGCCCGCCCCCGCGCCCCCCCC 902

Qy        381 GCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGG 440
Db        903 GCCCCCCCCCCCCCGCCCGCCCCCCCCCCCCCCCGCCCCCCGCCCCCCCCCCGCGCCCCC 962

Qy        441 GACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGAC 500
Db        963 CGCGCCCGCCCCGCGCGCGCCGCCCCGCCCCGCCCGCCCCCCGCGCCCCCCCCCCCGCC 1022

Qy        501 GAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCG 560
Db       1023 CCCCGCGCCCCCGCCCCCGCCCCCCCCGCCCGCGCCCCCGCCCCCCCCCGCCCCCCGCCC 1082

Qy        561 CCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTC 620
Db       1083 CCCCCCCCCCGCCCCGCGCCCCCCCCCCCGCCCCCGCGGCCCGGCCCGCCGCCCCCCCCG 1142

Qy        621 TTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAG 680
```

```
Qy        879 ATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTC 916
              |   ||   | | | | ||  | |  | ||  | |
Db       1141 GGCCCGGCGCCGCGGGCCCCCCGCCGCGNGCCCGCCNC 1178


RESULT 4
CG756569
LOCUS       CG756569               1406 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P051-4-B06.zb Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG756569
VERSION     CG756569.1  GI:37984265
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1406)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1406
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

  Query Match             11.9%;  Score 119.2;  DB 10;  Length 1406;
  Best Local Similarity   48.0%;  Pred. No. 4e-11;
  Matches  441;  Conservative    0;  Mismatches  473;  Indels    4;  Gaps    4;

Qy         85 CCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGC 144
              |||||| | ||  ||||  |  ||||   | |   | ||||   | | |   ||  ||
Db        362 CCCGCCGCCCCCGCGCCCCCCGGCGGGGCGGGGGCCGGCGGGCGGCGGCCCCCGCCCCCG 421

Qy        145 TCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCC 204
               | || |||||  | ||   ||   |  |  || |  | | |||| |   |  ||   ||
Db        422 GCGGC-CCGCGNGCCCCGCGCCGGGCGCGGCCCCGGCCCCCGCCCCCGGCCGCGGNGCCC 480

Qy        205 TGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGT 264
                  |  |  | ||| ||  | || ||  | ||   |    |||| ||   ||
Db        481 GCGCGCCCGCCCGNCCGGCCGGGGCCGCGCCCCGCCCGGGGCCGCGCGCGGCCGGCCCCG 540

Qy        265 CGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCG-GCCTCCGCGGCCCCGGCGGCC 323
              ||   ||| |    ||||||||| | |  || ||  || ||| || | ||||| ||  ||
Db        541 CGGGGGCCCCCCGGCCGGCGGCCGCGGGGGCGCCCCCGCGCCCCCCCCGCCCCCGCCCCC 600

Qy        324 TCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCG 383
               || |  | || || | | || |  ||    | | | ||||| ||||  | | | ||||
Db        601 GCCCCGCCGCCGGCCGGCCCCCCGCCCGGCCCCGGCCCCGCGCCCCCGCCCGGCGCCGCC 660

Qy        384 GCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGAC 443
               |  || |   | |  ||||||  | |  ||| |   |||       ||| |||  | | |
Db        661 CCGCCCCCCCCCCGCCCCGGCGCGGCCCCCGGCGCGCCCGCGGCCCGCGCCCGCCCGCGCC 720

Qy        444 GCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAG 503
              | ||| |||      |   | |   |  ||  |  |||| |||  |||||  |  |  | |
Db        721 CCGCGCCCCCCCCCCCCCCGGGGCCGCCGCCCCCCCGCGCCCCGGCCCCCCGGCCCCCCG 780

Qy        504 CTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCT 563
              | || |     |  |  |   | |      | || || ||  |  ||  | | | | |||
Db        781 CGCGGCCCGCGCCGCGGGGCGGCCGGGGCCCGGCGCCCGGCCCCGGCCCGCCGGGGGCCC 840

Qy        564 GGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTC 623
               ||   | |  || | |   ||||   |  ||| |    | | |   |  |  |
Db        841 CCGCCGCCCCCCCCCCCCGGCCCCCGCGCGCGCGCCCCGCCCGCCGCGCCCGGGCCCCCG 900

Qy        624 GCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATC 683
              ||    |   || | |||  | || |   ||    || | |  ||  ||  |  |  |
Db        901 CCCGCGCCCCCCGGCGGCCGCCCCGCCCGGCCCCCCGCCCCCGCCCGCGCGGCGCGGCGN 960

Qy        684 GCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGT 743
              |||||         || |  |    || ||     |||   |  ||   |  ||  ||
Db        961 NCGGCCGCCGCGCGCGGCGCCGCCGCGCCGCCCGCCCCGCCCGCCCCCNCCCCCGGCGGC 1020

Qy        744 GGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGG 803
              |  |   |  |||||||  |  |  |  |     | |||  | |  ||    |  ||
Db       1021 GCCCGNCCGCGCGGCGGCCCCCGCCGCGGCCCCCGGCCCCCCCCGGCCCCGCGCCCCGGC 1080

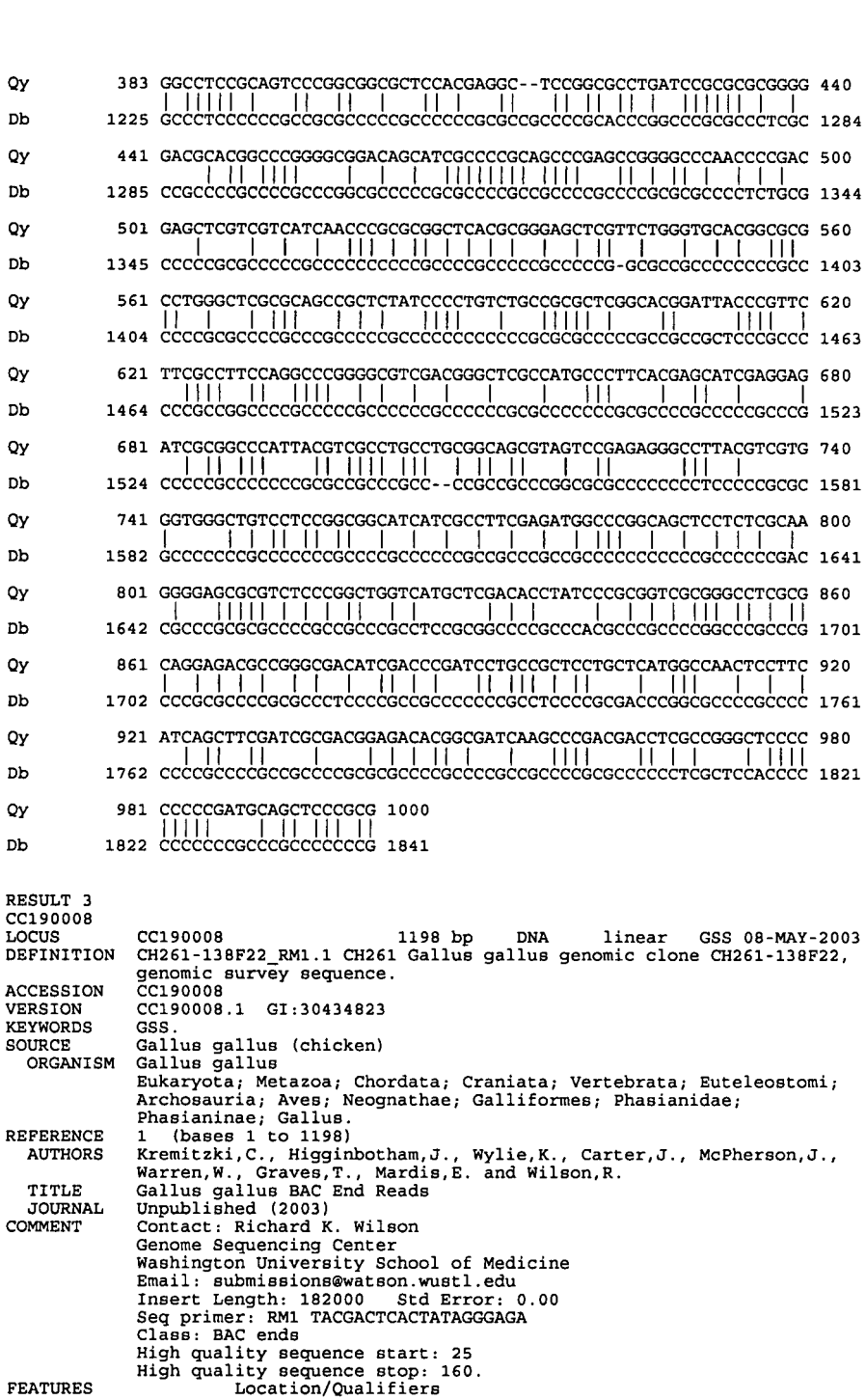
Qy        804 GAGCGCGTCTCCCG-GCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCA 862
              |  |||  | |||| || ||    || || | ||  | ||| | |||| ||| | | |
Db       1081 GGNCGCCCCCCCCGCGCCGGCGGCGCCCGCCCCCGGGCGCGCCGCGCGCGCCGCCCCCG 1140

Qy        863 GGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCAT 922
               | | |  || |||     |  ||    ||  |||| |  |  |   |    | ||  |
Db       1141 CGCGCCCGCGCGCGCGCCCCGCCGCGCCCGCCCGCCCGCGGGCCGCGGGCCCCCCCCCCG 1200

Qy        923 CAGCTTCGATCG-CGACGGAGACACGGCGATCAAGCCCGACGACCTCGCCGGGCTCCCCC 981
              | ||  ||  || || | | | | |||    |  ||||| |  |    || |  | |||
Db       1201 CCGCCCCGCGCGCCGNCCGCGCCCCGGGCGCCCCGCCCGGCCCCGNGCGCGCGGCCGCCC 1260

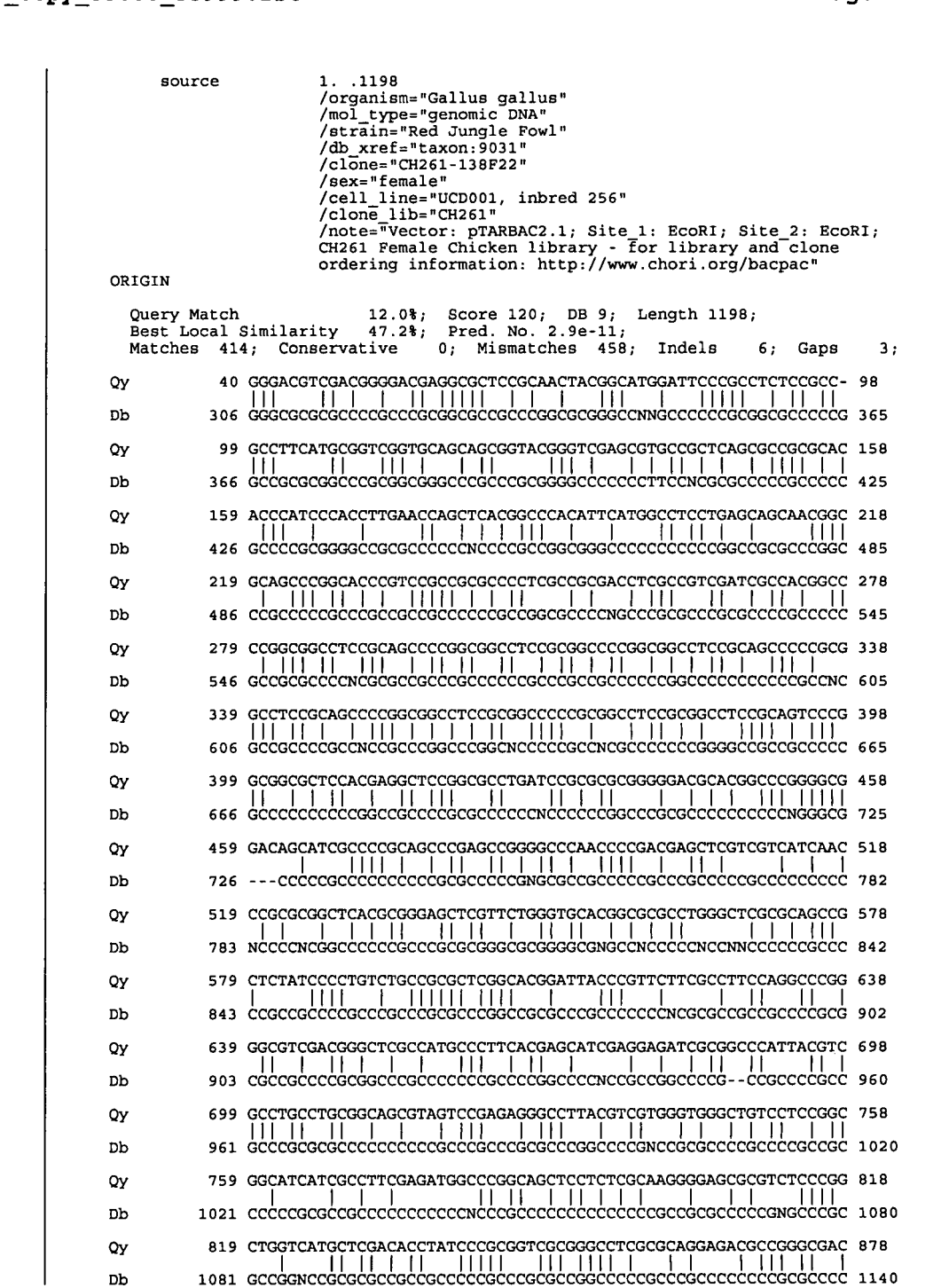
Qy        982 CCCCGATGCAGCTCCCGC 999
              | ||  ||  |  || |
Db       1261 GCGCGCCGCCCCCGCCCC 1278

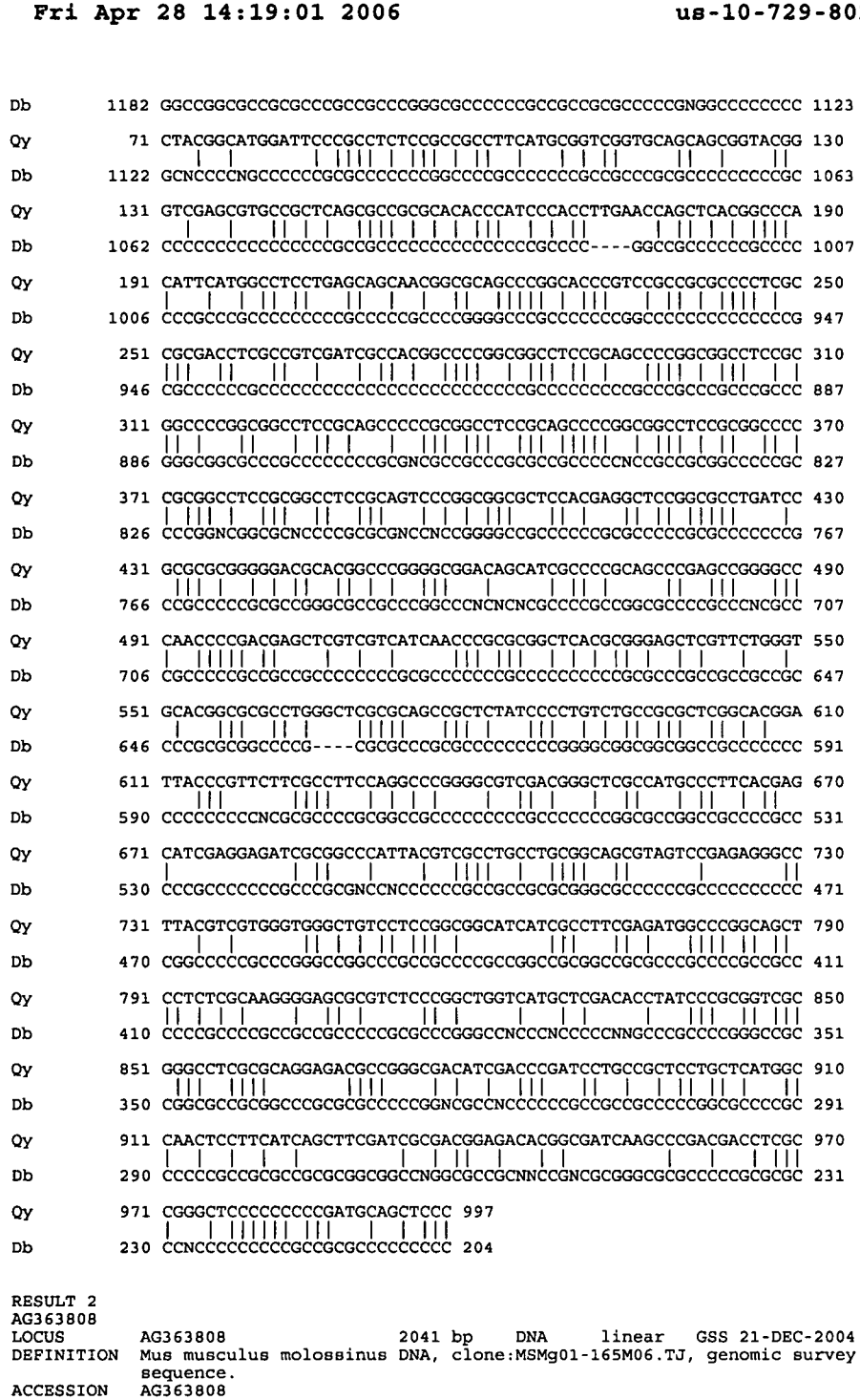

RESULT 5
AG072414
LOCUS       AG072414               996 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-063O11.R, genomic survey sequence.
ACCESSION   AG072414
VERSION     AG072414.1  GI:16624216
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 996)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
```



RESULT 3
CC190008
LOCUS CC190008 1198 bp DNA linear GSS 08-MAY-2003
DEFINITION CH261-138F22_RM1.1 CH261 Gallus gallus genomic clone CH261-138F22, genomic survey sequence.
ACCESSION CC190008
VERSION CC190008.1 GI:30434823
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1198)
AUTHORS Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: RM1 TACGACTCACTATAGGGAGA
Class: BAC ends
High quality sequence start: 25
High quality sequence stop: 160.
FEATURES Location/Qualifiers
source 1. .1198
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-138F22"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CH261 Female Chicken library - for library and clone ordering information: http://www.chori.org/bacpac"
ORIGIN

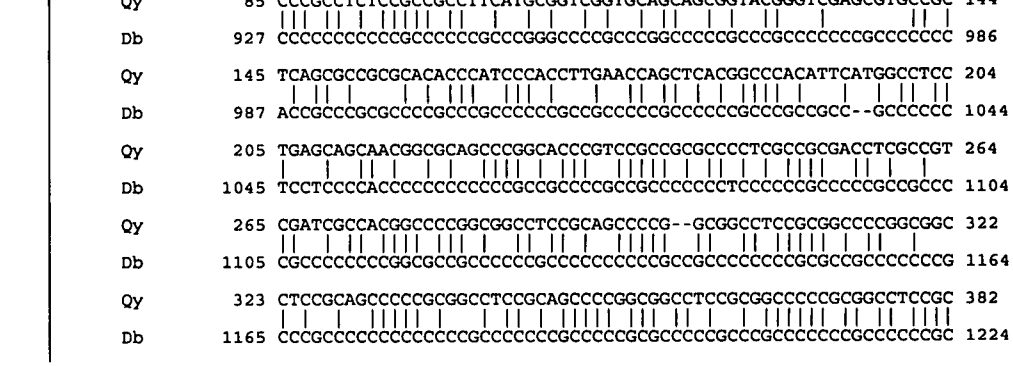
Query Match 12.0%; Score 120; DB 9; Length 1198;
Best Local Similarity 47.2%; Pred. No. 2.9e-11;
Matches 414; Conservative 0; Mismatches 458; Indels 6; Gaps 3;





RESULT 2
AG363808
LOCUS AG363808 2041 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-165M06.TJ, genomic survey sequence.
ACCESSION AG363808
VERSION AG363808.1 GI:47975013
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 2041)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .2041
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-165M06.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 12.1%; Score 120.8; DB 10; Length 2041;
Best Local Similarity 49.3%; Pred. No. 2e-11;
Matches 454; Conservative 0; Mismatches 457; Indels 9; Gaps 5;



GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 4956.14 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_55000_55999
Perfect score: 1000
Sequence: 1 tcacggcgatggtctgcgat.....cccccgatgcagctcccgcg 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	122.4	12.2	1628	10	CG757066	CG757066 P052-2-A0
	2	120.8	12.1	2041	10	AG363808	AG363808 Mus muscu
	3	120	12.0	1198	9	CC190008	CC190008 CH261-138
	4	119.2	11.9	1406	10	CG756569	CG756569 P051-4-B0
	5	119	11.9	996	10	AG072414	AG072414 Pan trogl
	6	118.6	11.9	1674	10	CL078342	CL078342 CH216-149
c	7	118.4	11.8	2017	10	AG435283	AG435283 Mus muscu
c	8	117	11.7	1692	10	AG396765	AG396765 Mus muscu
	9	117	11.7	1695	9	CC290874	CC290874 CH261-172
c	10	115.4	11.5	1425	10	AG441757	AG441757 Mus muscu
c	11	113.2	11.3	1523	10	AG448267	AG448267 Mus muscu
c	12	112.6	11.3	1286	10	AG448422	AG448422 Mus muscu
	13	112	11.2	1153	3	BM804596	BM804596 AGENCOURT
c	14	111.8	11.2	1046	5	BQ952554	BQ952554 AGENCOURT
c	15	111.6	11.2	1216	10	AG441574	AG441574 Mus muscu
	16	111.2	11.1	1392	10	AG390642	AG390642 Mus muscu
c	17	111.2	11.1	1961	10	AG435522	AG435522 Mus muscu
c	18	110.6	11.1	1328	10	AG441962	AG441962 Mus muscu
c	19	110.4	11.0	916	10	CW947986	CW947986 TcB31.2_G
c	20	110.4	11.0	1821	10	CL090560	CL090560 ISB1-17N1
	21	110	11.0	1235	10	CL081604	CL081604 CH216-163
c	22	110	11.0	1450	10	AG429738	AG429738 Mus muscu
c	23	109.6	11.0	1453	8	DN806685	DN806685 76982446
c	24	109.6	11.0	1713	10	AG381912	AG381912 Mus muscu
	25	109.4	10.9	897	10	AG046101	AG046101 Pan trogl
c	26	109.4	10.9	1177	10	AG441911	AG441911 Mus muscu
c	27	109.2	10.9	1839	10	CG747711	CG747711 P041-2-E0
c	28	109	10.9	1506	10	AG442117	AG442117 Mus muscu
c	29	109	10.9	1675	10	AG360980	AG360980 Mus muscu
	30	108.8	10.9	1598	10	CL078355	CL078355 CH216-149
c	31	108.6	10.9	1324	10	AG448157	AG448157 Mus muscu
c	32	108.6	10.9	1406	10	AG441598	AG441598 Mus muscu
	33	108.2	10.8	1004	6	CD051228	CD051228 AGENCOURT
	34	107.6	10.8	1142	10	CL461680	CL461680 SAIL_1150
	35	107.6	10.8	1471	10	CG748176	CG748176 P042-1-A0
c	36	107.6	10.8	1552	10	AG430101	AG430101 Mus muscu
	37	107.4	10.7	961	10	AG592904	AG592904 Mus muscu
	38	107.2	10.7	1272	10	AG365674	AG365674 Mus muscu
	39	107	10.7	1101	10	AG039543	AG039543 Pan trogl
	40	107	10.7	1542	10	AG032943	AG032943 Pan trogl
	41	106.8	10.7	776	10	CNS010RY	AL099352 Drosophil
c	42	106.8	10.7	1016	11	CNS03LU9	AL249930 Tetraodon
c	43	106.8	10.7	1822	10	AG435170	AG435170 Mus muscu
	44	106.4	10.6	1101	10	CNS0153F	AL104949 Drosophil
c	45	106.4	10.6	1275	10	AG442214	AG442214 Mus muscu

ALIGNMENTS

RESULT 1
CG757066/c
LOCUS CG757066 1628 bp DNA linear GSS 24-OCT-2003
DEFINITION P052-2-A02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG757066
VERSION CG757066.1 GI:37985257
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1628)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1628
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 12.2%; Score 122.4; DB 10; Length 1628;
Best Local Similarity 45.7%; Pred. No. 1.1e-11;
Matches 451; Conservative 0; Mismatches 528; Indels 8; Gaps 2;

Qy 11 GGTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGGCGCTCCGCAA 70

Search completed: April 27, 2006, 15:53:00
Job time : 628.947 secs

```
Db        870 TCCGGCCCTGCCGCCGGCCCCGCCGTCACCCCCGATAGCGACCGCGGCCCCGCCGTTGCC 811

Qy        199 GCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCG-TCCGCCGCGCCCCTCGCCGCGACC 257

Db        810 GCCGACGCCGCCGGCACCGTAGAGCATGCCGCCCGCGCCGCCGGACCCACCGTCACCCCC 751

Qy        258 TCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCG 317

Db        750 GATACTTGCGCTGTTGCTGGCGCCCCCGCCATTGCCGCCGGCGCCGCCGGCGCCGCCCGC 691

Qy        318 GCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCG-CGGC 376

Db        690 GCCGCCGCCACCACCCAACAGCCCACCGGTGCCGCCCGCGCCGCCGCTACCCCCGATGGC 631

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436

Db        630 GTCGCCGGCCCCGCCGGCGCCGCCGCCGCTGAACACGCCGCCGGCGCCACCAGCCCCGCC 571

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496

Db        570 GTCACCCCCGATCGCCACCGCGGACCCGCAGCCCCCGCCGGCGCCGCCGATGCCGAAGAG 511

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556

Db        510 CCCGCCGGCGCCGCCCGCCCCGCCGACCCCGCCGCGGCCGGTCCCGGCTAAGTCGCTGGC 451

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616

Db        450 GCCGCCGGTCCCGCCGTTGCCGATCAAGATGCCGCCGGCCCCGCCGGCCCCCCCGCTCCC 391

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676

Db        390 CGGCGCCGCATCGGTGCCGTCGCCGATCAACGGGCGTCCCAACAGCGCCAGGGTGGGTGC 331

Qy        677 GGAGATCGCG 686

Db        330 ATTGATGGCG 321


RESULT 14
US-10-775-169-313
; Sequence 313, Application US/10775169
; Publication No. US20050287532A9
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Burczynski, Michael
;  APPLICANT:  Twine, Natalie
;  APPLICANT:  Dorner, Andrew
;  APPLICANT:  Trepicchio, William
;  TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
;  FILE REFERENCE: AM101080 (031896-013000)
;  CURRENT APPLICATION NUMBER: US/10/775,169
;  CURRENT FILING DATE:  2004-02-11
;  NUMBER OF SEQ ID NOS: 5278
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 313
;   LENGTH: 1511
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-169-313

  Query Match             7.6%;  Score 76.4;  DB 8;  Length 1511;
  Best Local Similarity   48.6%;  Pred. No. 3.7e-06;
  Matches  243;  Conservative    0;  Mismatches  251;  Indels    6;  Gaps    1;

Qy         31 TGCAGGCCAGGGACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCC 90

Db        490 TGCAGTGGACCGAAGTGGACAGCGACTCGGTGCCCGCGTTCTACCGCGAGACGGTGCAGG 549

Qy         91 TCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCG 150

Db        550 TGGGGCGCTGCCGCCTGCTGCTGGCGCCGCGGCCCGTCGCGGTCGCGGTGGCTGTCAGCC 609

Qy        151 C------CGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCC 204

Db        610 CGCCCCTCGAGCCGGCCGCTGAGTCCCTCGACGGCCTCGAGGAGGCGCCGGAGCAGCTGC 669

Qy        205 TGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGT 264

Db        670 CTAGTGTCCCGGTCCCGGCCCCGGCGTCCACCCCGCCCCCAGTCCCGGTCCTGGCTCCAG 729

Qy        265 CGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCT 324

Db        730 CCCCGGCCCCGGCTCCGGCTCCGGTCGCGGCTCCGGTCGCGGCTCCGGTCGCGGTCGCGG 789

Qy        325 CCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGG 384

Db        790 TCCTGGCCCCGGCCCCGGCCCCGGCCCCGGCTCCGGCTCCGGCCCCGGCTCCAGTCGCGG 849

Qy        385 CCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACG 444

Db        850 CCCCGGCCCCAGCCCCGGCCCCGGCCCCGGCCCCGGCCCCCGCCCCGGCCCCGGCCCCGG 909

Qy        445 CACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGC 504

Db        910 ACGCGGCGCCTCAAGAGAGCGCCGAGCAGGGCGCGAACCAGGGGCAGCGCGGCCAGGAGC 969

Qy        505 TCGTCGTCATCAACCCGCGC 524

Db        970 CTCTCGCTGACCAGCTGCAC 989


RESULT 15
US-11-052-554A-530
; Sequence 530, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
;  APPLICANT: Sachdeva, et al.
;  TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
;  TITLE OF INVENTION:  PROTEINS OF THERAPEUTIC POTENTIAL
;  FILE REFERENCE: 30853/40359A
;  CURRENT APPLICATION NUMBER: US/11/052,554A
;  CURRENT FILING DATE:  2005-02-07
;  PRIOR APPLICATION NUMBER: US 60/589,227
;  PRIOR FILING DATE: 2004-07-20
;  PRIOR APPLICATION NUMBER: IN 173/DEL/2004
;  PRIOR FILING DATE: 2004-02-06
;  NUMBER OF SEQ ID NOS: 763
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 530
;   LENGTH: 4617
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-530

  Query Match             7.6%;  Score 76;  DB 14;  Length 4617;
  Best Local Similarity   58.3%;  Pred. No. 3.4e-06;
  Matches  133;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        232 CGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCG 291

Db       3465 CGCCGGCGGCAACGGCGGCTACGGCGGTCCGGGAGGCCCCGAAGGCAATGGCGGCCTCGG 3524

Qy        292 CAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCC 351

Db       3525 CGGCAACGGCGGCGCCGGCGGCAACGGCGGCGTCAGCACCACGGGCGGCGACGGCGGCGC 3584

Qy        352 CGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACG 411

Db       3585 CGGCGGCAAGGGCGGCAACGGCGGCGACGGCGGGAACGTCGGTTTGGGCGGTGACGCCGG 3644

Qy        412 AGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGG 459

Db       3645 CTCCGGCGGCGCGGGCGGCAATGGCGGTATCGGCACCGACGCGGGCGG 3692
```

```
Db      48582 CCGCCGACGGCCCGGCCGAGCCGCCGCGCGCGGACCCCACGCCGCCGCCCCCGCCAGCGC 48523

Qy        177 CAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCC 236
              || |   |  |   |     ||| || |    ||    ||||  | |||| |||  ||
Db      48522 CACCCACCACCGGCCGCGGCGCGCCGCCCGCCGGGCGCGCGCGCGCCGCGGCCCCCCGCC 48463

Qy        237 GCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCC 296
              | |||||| |    |||| || |   |  |      |  |   |||   | || |||| ||
Db      48462 GGCGCGCCGCGGCACGCGCCCCCCACCACCCCGCCGCCCCGCCCCCGGCCCACCGCGGCA 48403

Qy        297 CCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAG--CCCCCGCGGCCTCCGCAGCCCCGG 354
              ||| |  || ||||||  ||  ||||| ||| ||  | ||||||  || |   |  | |||
Db      48402 CCGCCCCCCCCCCGCGGGGCCCCCGGCCCCCGAAGACCGCCCGCGCCCCCGCCGCCGCCGC 48343

Qy        355 CGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGG 414
              |||  ||| ||||||  | | | ||   | | || ||  | |  | || | || |  |
Db      48342 GCGCCCGCGCCGCCCCCCCCGGCCCCCGCGACGCCCCACCCACCCCCGCCCCCCGCCCGC 48283

Qy        415 CTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGC 474
              ||  || |      | | ||    |  | | || |||  | |   |     |||||
Db      48282 ACCCCCCGGCCAGCACCCCCGACCCGCGCCCCCGCCCCCCGCCCCCCCCGCACGCCCACG 48223

Qy        475 AGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCG 534
              |    |||   |||  |||| |   | |  |  |  |  |||| ||  |  || ||
Db      48222 CCGCGCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCGGGCCCCCGACCCG 48163

Qy        535 GGAGCTC 541
              |  | |
Db      48162 CGGCCCC 48156


RESULT 12
US-10-496-351-1
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             7.7%;  Score 76.6;  DB 7;  Length 14055;
  Best Local Similarity   47.3%;  Pred. No. 2e-06;
  Matches  398;  Conservative    0;  Mismatches  429;  Indels   14;  Gaps    5;

Qy         96 GCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCG 155
              ||||||  |   || || |    || ||     |  || | ||     |  |  |   | |
Db       8284 GCCGCCCGCCGCCGCTCCGCCGGGCTGCCCGCCACCTCCATCGCCTGGGGCCACTGGGAG 8343

Qy        156 CACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAAC 215
              ||||||    ||  |      | |||||| |||  |   |  |  | |  |   |||  ||
Db       8344 CACACCAGCGCCCTCACCGCCGGGCTCACCGCCGCCGACCTCGCGCGCGGGGCCAGGCAC 8403

Qy        216 GGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACG 275
              |||||  |  || | |  || ||| | | || |||  |   |     |  || ||||| ||
Db       8404 GGCGCGGGCTCGCTCACCACCACCGAGGCGCTGGCCCTGCTCGACGCCGCGCTCGCCTCG 8463

Qy        276 GC----CCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGC 331
               |    |||| || ||||  |  | |   || || |||| || ||| ||   |||
Db       8464 CCGGAACCCGCCGTCCTCGCCGCCGCGCTCGACCCCCGCCGCGCCGCCGAGGTCCCGCCG 8523

Qy        332 CCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGC 391
```

```
              |  |   ||  || | |||    ||  |   |||| | | ||     |||  |||  |||
Db       8524 CTGCTGCGCGACCTC-GCCGGACCGCGCCGGGCGGTCGCGGCCCAGGCCGTCGCCGGCGC 8582

Qy        392 AGTCCCGGCGGCGCTCCACGAGGCTCCGGCG----CCTGATCCGCGCGCGGGGGACGCAC 447
              | || | ||  | || ||| ||    ||||    ||    ||| || ||||   |||
Db       8583 GGACCGGCCGCTGATCGACGTGGTCAGGGCGCACACCGCGACCGTGCTCGGGCACCGCGA 8642

Qy        448 GGC--CCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCT 505
              ||  |||  | ||||   | |  | ||       ||||  |||   | ||  |   |
Db       8643 CGCCTCCGCCGTGGACGCGAACCGCGCGTTCAAGGAGCTGGGCGTCGACTCGCTCACGGC 8702

Qy        506 CGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGG 565
              ||| |   |   |   ||   | |   || || |  | ||  | |   || | |  | |
Db       8703 CGTGGAGCTGCGCAACCGGCTCGCCGCCGCGACCGGGCTCGCGCTCGCCGCCACCTCGGT 8762

Qy        566 GCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGC 625
              | |||  || |||  |     | |||     ||||||  |       |   | |       |||
Db       8763 GTTCGACCACCCGACCCCGGCGCTGCTCACCGCGCACCTGCTCGCCGAGCTCGGGCACGC 8822

Qy        626 CTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGC 685
              |  |   ||| ||  |  ||      | |||||   ||  |   | |  ||| | |   |
Db       8823 CCCCGCCGCCGGGACCACCGCGCCCCCCGCCACCGCCGCCGTCACCGCCGACGCGGGGGA 8882

Qy        686 GGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGG 745
              | ||       |||  |   ||      ||    ||  | |     | |  |  ||  |
Db       8883 GCCCATCGCGATCGTGTCGATGGCCTGCCGCTTCCCCGGCGACGTGTCCTCCCCGGAGGA 8942

Qy        746 GCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGA 805
              ||||   |  || |    ||  |   ||   |  |   | | ||| | |  |  |||
Db       8943 CCTGTGGGACCTCGTCGCGGGCGGCGGGGACGCGCTCACCCCGTTCCCCGCCGACCGGG- 9001

Qy        806 GCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGA 865
              ||      || |  || | || ||| | || | || |     ||| || | |   || |
Db       9002 --GCTGGGACCTGGCGGCCCTGTTCGCCCCCGACCCGGACCGCGCCGGGCGCAGCCACGC 9059

Qy        866 GACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAG 925
                  |||||   ||  | |   | |  ||  | |   |   | |   | |  ||||||
Db       9060 CCGCGAGGGCGGGTTCCTCGCCGACGTCGCGGGCTTCGACGCCGACCTGTTCGGCATCAG 9119

Qy        926 C 926
              |
Db       9120 C 9120


RESULT 13
US-11-052-554A-553/c
; Sequence 553, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
;  APPLICANT: Sachdeva, et al.
;  TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
;  TITLE OF INVENTION:  PROTEINS OF THERAPEUTIC POTENTIAL
;  FILE REFERENCE: 30853/40359A
;  CURRENT APPLICATION NUMBER: US/11/052,554A
;  CURRENT FILING DATE:  2005-02-07
;  PRIOR APPLICATION NUMBER: US 60/589,227
;  PRIOR FILING DATE: 2004-07-20
;  PRIOR APPLICATION NUMBER: IN 173/DEL/2004
;  PRIOR FILING DATE: 2004-02-06
;  NUMBER OF SEQ ID NOS: 763
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 553
;   LENGTH: 1086
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-553

  Query Match             7.6%;  Score 76.4;  DB 14;  Length 1086;
  Best Local Similarity   48.5%;  Pred. No. 4e-06;
  Matches  267;  Conservative    0;  Mismatches  281;  Indels    2;  Gaps    2;

Qy        139 TGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATG 198
              | | || | ||  |  || |  || || | ||    | |  |    | ||| |  |
```

```
Qy        442 ACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACG 501
Db        773 GCCAGGGCGTCCGGCGCGACCAGCGCGCCCGCCTGCACCGCCACACCCAGCGCCTGGAGA 832

Qy        502 AGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGC 561
Db        833 TGCGGCACATCCAGCGCGAGCGGCCCAGCCGCGTCGCGGGCGTGCTCGCCGACGACGTCG 892

Qy        562 CTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCT 621
Db        893 ACCAGCTCGGCGCGTTCTTCTCCGGCCCGGCGGGCGAGGCGGTCCAGCTCGCCACCAGCG 952

Qy        622 TCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGA 681
Db        953 CGCTGATCCTGGTCCCGTCGTTCCTGGTGCTCGTGCCCGGCATCGCCTGGATCGCGTTCC 1012

Qy        682 TCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGG 741
Db       1013 TGCCCCTGCCGTTCGTCGCGGCCCTGTCCCTGC---GCCAGCGCGACCGCTCCGCCGCCG 1069

Qy        742 GTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAG 801
Db       1070 ACTACGCCAGCGCGGGCGAGCAGGCCGCCCTGCTGCGCAGCAGGCTCTCCGCCGCACTGG 1129

Qy        802 GGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGC 861
Db       1130 AGGCCGCCGCCACGGTCAAGAGCGCCTGCGCCGAGGACCACGAGGCCGAGCTGGTCGACC 1189

Qy        862 AGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCA 921
Db       1190 TGCTGGAGCTGGCGGGCGCCGACGCGGCCACGCGCACCCAGCAGCACGCCGTCCGGTTCG 1249

Qy        922 TCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCGACGACCTCGCCGGGCTCCCCC 981
Db       1250 GCGAGGGCATCCGCGCCTGGACCACCTGCTCCATGGCGGGCACCCTGCTCGTGGGCGGGC 1309

Qy        982 CCCCGATGC 990
Db       1310 TCAAGGTGC 1318
```

RESULT 10
US-11-052-554A-529/c
; Sequence 529, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 529
; LENGTH: 3240
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-529

Query Match 7.7%; Score 77.2; DB 14; Length 3240;
Best Local Similarity 47.0%; Pred. No. 2.2e-06;
Matches 311; Conservative 0; Mismatches 343; Indels 8; Gaps 2;

```
Qy         45 GTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTC 104
Db       1749 GCCACCGCCGACGGTGCGGTCGCCGCCGGTGCCGCCGGCACCGCCGTCGCCGCCGGCGCC 1690

Qy        105 ATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCAT 164
Db       1689 ACCCTGACCCCCGGTCCCGCCACCGGTGCCGCCTTGGCCGCCGGCGCCACCCTTTCCGGC 1630

Qy        165 CCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCC 224
Db       1629 GGCGCCGCCGTTACCGGCGGCCCCCGCGGTGCCCGCGGTGCCGCCGGTGATATTGCCGCC 1570

Qy        225 CGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCG 284
Db       1569 GTTGCCGCCGGTCCCGCCGGCACCGCCGGCACCACCTTGACCACCGGCACCGCCGCCGGC 1510

Qy        285 GCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCC 344
Db       1509 GCCGCCCTGGCCACCTTGTCCGCCGTCGCCGGCCGCCCCGCCGGTGCCGCCGGCGCCTCC 1450

Qy        345 GCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCG-------CAGTCCC 397
Db       1449 GGTGCCGCCGGCGCCGCCGCTGCCGCCGTCGCCTCCGGTGCCGCCGGCGCCGCCGCTGCC 1390

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
Db       1389 GCCGTTGGTAGCTGAGGAGGCTGCGCCGGCGCCGCCGGTGCCGCCGGCACCGCCTTGGCC 1330

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
Db       1329 GCCGTTGCCAGCGGTGCCGTTAGCGCC-GCCACCGGCGCCGCCAGCCCCGGCCGCCCCAC 1271

Qy        518 CCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCC 577
Db       1270 CGGCACCACCGCTGCCGCCTTGACCTCCTTGCCCGCCGTTGCCGCCGGGACTGGTGTTGG 1211

Qy        578 GCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCG 637
Db       1210 GCCCGCTGCCGCCGGCGCCGCCCGCACCGCCGACTTCGCCGTCCGCGCCGTTGCCGCCCT 1151

Qy        638 GGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGT 697
Db       1150 TGCCGCCGCCGACGCCACCGATACCGTCGCCGCCGGCGCCACCACTGCCGCCTCTGCCGC 1091

Qy        698 CG 699
Db       1090 CG 1089
```

RESULT 11
US-10-330-773-26/c
; Sequence 26, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 687411
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(687411)
; OTHER INFORMATION: n = A,T,C or G
US-10-330-773-26

Query Match 7.7%; Score 77; DB 9; Length 687411;
Best Local Similarity 50.4%; Pred. No. 6.7e-07;
Matches 215; Conservative 0; Mismatches 210; Indels 2; Gaps 1;

```
Qy        117 CAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAAC 176
```

```
Db        112 CGGCCCCGCCGTCGCCGAGCAGCCACCCACCGGGGGCCCCGGTGGCCCCGCTGCCGACCG 53

Qy        530 ACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGC 567
Db         52 CCCCGGGGGTGCCGTTGCCGATCAGCGGTCGCCCGGTC 15


RESULT 8
US-10-478-943D-1
; Sequence 1, Application US/10478943D
; Publication No. US20060073574A1
; GENERAL INFORMATION:
;  APPLICANT: University College Dublin, National University of Ireland
;  TITLE OF INVENTION: Engineered biosynthesis of novel polyenes
;  FILE REFERENCE: PC-1636US
;  CURRENT APPLICATION NUMBER: US/10/478,943D
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: S2001/0527
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 113193
;   TYPE: DNA
;   ORGANISM: Streptomyces nodosus
US-10-478-943D-1

  Query Match             7.8%;  Score 78;  DB 7;  Length 113193;
  Best Local Similarity   44.7%;  Pred. No. 6.6e-07;
  Matches  389;  Conservative    0;  Mismatches  475;  Indels    6;  Gaps    2;

Qy         18 GATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGGCGCTCCGCAACTACGGC 77
Db      22997 GACGCCGTCGCCGTCCGCGCCAGCGGCGTCTTCCTGCGCCGGCTCGCCCACAGCCCCGCC 23056

Qy         78 ATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGC 137
Db      23057 GCCGACACCGTCGGCTCCGCCTTCGACCCGGCGGCCGGCACCGTCCTGATCACCGGCGGC 23116

Qy        138 GTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCAT 197
Db      23117 ACCGGCGGCATCGGCGGACACCTCGCCCGCCGCCTCGCCCGCGACGGCGCCGCCCATCTG 23176

Qy        198 GGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACC 257
Db      23177 CTGCTCACCAGCCGCCGCGGCCCCGACGCCCCCGGCGCCGGCGAACTCCGCGCCGAACTG 23236

Qy        258 TCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCG 317
Db      23237 GAGGAGTCGGGCGCCCGTGTCACGATCGCCGCCTGCGACGCCGCCGACCGCGACGCGCTC 23296

Qy        318 GCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCC 377
Db      23297 GCCGCACTCCTCGCCACGGTGC---CCGAGGACGCGCCCCTCACCGCCGTCTTCCACACC 23353

Qy        378 TCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCG 437
Db      23354 GCGGGCGTCGTCGACGACCATGTGGTGGACGAGCTGACCCCGGAGAGCTTCGCCACCGTC 23413

Qy        438 GGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCC 497
Db      23414 CTGCATGCCAAGACGGTCGCCGCCCGCCATCTGCACGACCTCACCGCCGGCCACGACCTC 23473

Qy        498 GACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGC 557
Db      23474 GCCGCGTTCGTCCTGTTCTCCTCGACCGCCGGCCTGATCGGCGCCGCCGGACAGGGCAAC 23533

Qy        558 GCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCG 617
Db      23534 TACGCCGCCGCCAACTCCCAGCTCGACGCCCTCGCCGAGCAGCGCCGCGCCCAGGGCCTG 23593

Qy        618 TTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAG 677
```

```
Db      23594 CCCGCCCTGTCCGTGGCCTGGGGACCCTGGGCCGGCTCCGGCATGGTCGCCGACGCCGCC 23653

Qy        678 GAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTC 737
Db      23654 GAGATCGCCGCCCGGGTGCGCCGCGGCGGATTCGAACCGCTCGACCCCGCCTCCGGTGTC 23713

Qy        738 GTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCG 797
Db      23714 GCCGCGCTGCTGCGCGCCGTCGACAACGGCGACACCTCCGTCGCCGTCGCCGACATC--- 23770

Qy        798 CAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTC 857
Db      23771 GACTGGGAACGCTTCCTGCGCGCCTTCGCGTCGGCCCGCCCGCTGCCGCTCGTCGCCGAC 23830

Qy        858 GCGCAGGAGACGCCGGGCGACATCGACCCG 887
Db      23831 CTCCCGGAGACGGCGGGCGCCAACGCCCCG 23860


RESULT 9
US-10-496-351-8
; Sequence 8, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 2247
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-8

  Query Match             7.8%;  Score 77.8;  DB 7;  Length 2247;
  Best Local Similarity   44.0%;  Pred. No. 1.8e-06;
  Matches  426;  Conservative    0;  Mismatches  537;  Indels    6;  Gaps    2;

Qy         25 TCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATT 84
Db        353 TCGCGCTCCGGCCCGCGAGCGCCGTCCTGCGCCGCCCGCTCGCCCTGGTCGGCGTGGTCA 412

Qy         85 CCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGC 144
Db        413 CCGGCGTCTCGTCCGCCGTCCTGCGCCACGCCTGGCGCGCCCTCCTGCGGCCCCTGCGCC 472

Qy        145 TCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACG---GCCCACATTCATGGCC 201
Db        473 GCACCGCCGCGGGCACCCGGCTCTCCGCCCACCCGCTGGTGCGCGCCCTCGGCCGCCGCC 532

Qy        202 TCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGC 261
Db        533 GGTCGGCGCTCGCCGCCGCCGCCGCGCTGACGGTGGCCTGCCAGCTCGCCGAGCTCGGCC 592

Qy        262 CGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGG 321
Db        593 TCGGGCTGTTCCTCGGCTGGATCCTGCTGGTGCTGATCAAGGGCGAGTACGCCCCGCTCA 652

Qy        322 CCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCG 381
Db        653 CCGCGCTGGGCGTCACGAGCGCCCCCGCCCAGCTGCTCGCGCTCGCCGCCGGGGCCGCCG 712

Qy        382 CGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGG 441
Db        713 CGACCTGCGGGGTCGTCGTCGTGCTCTCCGCCGCCTCCGGCGCCCGCTGGCGCGCGCTCG 772
```

```
Qy        329 AGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTC 388
              ||  | |||||| | || ||| | || ||| |||  ||  ||||| || | ||  |  |
Db        460 GGCTGCTGCGGCCGCTGCCGCCGCTGCTGCCGCCGTCGCTGCCGCGCCCCCGGCCCCTGC 519

Qy        389 CGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACG 448
              ||| | | |  ||| | |  | ||| | |  ||   ||||     |        |||
Db        520 CGCCGCCTCTACGGTGGACACAGCGGCCCTGAAGCAGCCTCCGGCGCCCCCTCCGCCACC 579

Qy        449 GCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGT 508
              ||    | |        |||  |  || ||| |||     ||  | ||| |    |||
Db        580 CCCGCCAGTGTCGGCGCCCGCGGCCGAGGCCGCGCCCCCCGCCTCCGCCGCCACTATCGC 639

Qy        509 CG 510
              ||
Db        640 CG 641
```

RESULT 6
US-10-802-796-4/c
; Sequence 4, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 4
; LENGTH: 1280
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-802-796-4

Query Match 7.8%; Score 78.4; DB 8; Length 1280;
Best Local Similarity 48.5%; Pred. No. 1.6e-06;
Matches 251; Conservative 0; Mismatches 261; Indels 6; Gaps 1;

```
Qy         50 CGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCG 109
              ||  | |   ||  ||| || |   | |  || || |  | || ||||||  || |    |
Db        526 CGCCGGCCCCGCCGTCCCCACCGGTGTCCAGGTTTTCGCCGTCACCGCCGGCTCCGCCGG 467

Qy        110 GTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCAC 169
              | |  | ||   |  |    ||  || | ||   | || ||||  || ||      | |
Db        466 CGCCGCCCGGCCCGGCGAGCATTCCGGCATTCCCGCCGGCCCCGCCAACCCCGCCGTCGC 407

Qy        170 CTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCA 229
              | | |         | |||         |      | ||| ||       || | ||
Db        406 CATTAGTGCTGAGCCCGCCGGTCCCGCCGTGACCTCCGCCGGCACCGATCAGCCCGGCCA 347

Qy        230 CCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTC 289
              | ||||||||  |||  || | | |||   ||  |  |||||||   |  ||| | ||
Db        346 GCAGTCCGCCGGTCCCGCCGGCCCCACCGGCCCCGCCGTCGCCACCAACAAGGCCGGTTC 287

Qy        290 CGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGC 349
              | | | ||| || | | || | | |      || | | ||||||  || |||
Db        286 CACCGGCCCCGCCGGCCCCACCGGCGCCCCACAGCCCACCGACCCCGCCTCCCCCG---- 231

Qy        350 CCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCA 409
                | |  ||| ||| |  |   || | ||||||  |   |||| | ||| || |    ||
```

```
Db        230 --CCGGTGCCGCCGAGTACGGTGCTGGCTCCGCCGATCCCGCCGGCCCCGCCGGTACCCC 173

Qy        410 CGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGC 469
              || |       || |  |  | ||| |||  ||| ||  || |    | | |   |||
Db        172 ACAGCCCGGCAGCCCCGCCCGCCCCGCCGGGCGCGCCCGAGCCCGCCGCGCCGGACCCGC 113

Qy        470 CCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTC 529
              |   |   |||    | | ||      |||  || |  |  ||     |  | |   || |
Db        112 CGGCCCCGCCGTCGCCGAGCAGCCACCCACCGGGGGCCCCGGTGGCCCCGCTGCCGACCG 53

Qy        530 ACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGC 567
              | |||| |  |  |   |  ||  |  |||| || |
Db         52 CCCCGGGGGTGCCGTTGCCGATCAGCGGTCGCCCGGTC 15
```

RESULT 7
US-11-221-284-4/c
; Sequence 4, Application US/11221284
; Publication No. US20060063182A1
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/11/221,284
; CURRENT FILING DATE: 2005-09-08
; PRIOR APPLICATION NUMBER: US/09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1280
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-11-221-284-4

Query Match 7.8%; Score 78.4; DB 11; Length 1280;
Best Local Similarity 48.5%; Pred. No. 1.6e-06;
Matches 251; Conservative 0; Mismatches 261; Indels 6; Gaps 1;

```
Qy         50 CGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCG 109
              ||  | |   ||  ||| || |   | |  || || |  | || ||||||  || |    |
Db        526 CGCCGGCCCCGCCGTCCCCACCGGTGTCCAGGTTTTCGCCGTCACCGCCGGCTCCGCCGG 467

Qy        110 GTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCAC 169
              | |  | ||   |  |    ||  || | ||   | || ||||  || ||      | |
Db        466 CGCCGCCCGGCCCGGCGAGCATTCCGGCATTCCCGCCGGCCCCGCCAACCCCGCCGTCGC 407

Qy        170 CTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCA 229
              | | |         | |||         |      | ||| ||       || | ||
Db        406 CATTAGTGCTGAGCCCGCCGGTCCCGCCGTGACCTCCGCCGGCACCGATCAGCCCGGCCA 347

Qy        230 CCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTC 289
              | ||||||||  |||  || | | |||   ||  |  |||||||   |  ||| | ||
Db        346 GCAGTCCGCCGGTCCCGCCGGCCCCACCGGCCCCGCCGTCGCCACCAACAAGGCCGGTTC 287

Qy        290 CGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGC 349
              | | | ||| || | | || | | |      || | | ||||||  || |||
Db        286 CACCGGCCCCGCCGGCCCCACCGGCGCCCCACAGCCCACCGACCCCGCCTCCCCCG---- 231

Qy        350 CCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCA 409
              | |  ||| ||| |  |   || | ||||||  |   |||| | ||| || |    ||
Db        230 --CCGGTGCCGCCGAGTACGGTGCTGGCTCCGCCGATCCCGCCGGCCCCGCCGGTACCCC 173

Qy        410 CGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGC 469
              || |       || |  |  | ||| |||  ||| ||  || |    | | |   |||
Db        172 ACAGCCCGGCAGCCCCGCCCGCCCCGCCGGGCGCGCCCGAGCCCGCCGCGCCGGACCCGC 113

Qy        470 CCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTC 529
```

```
Db      68398 CGCCCGCGTCACCACCGCCGCCTGCGACGTCGCCGACCGCGCCGCGCTCGCCGCCCTGCT 68339

Qy        329 AGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTC 388

Db      68338 GACCGGCCTGGAGCGCGCGGGCGAGCAGGTCACCGCCGTCGTGCACGCGGCGGGCGCCAA 68279

Qy        389 CGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACG 448

Db      68278 CGCCCAGACCCCCGTCGCCGACACCACCCCCGAGGAGCACGCGCGCGTCCAGGCCGCCAA 68219

Qy        449 GCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGT 508

Db      68218 GGCGCTCGGCGCCGAGCACCTCGACGAGCTGCTGGGCGGGCGCCCGCTCGACGCGTTCGT 68159

Qy        509 CGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCT 568

Db      68158 CCTGTTCTCCTCCAACGCGGGCGTGTGGGGCAGCGGCGGCCAGTGCGCCTACGCCGCCGC 68099

Qy        569 CGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTT 628

Db      68098 CAACGCCCGCCTCGACGCACTGGCCCAGCGCCGCCGCGCGCACGGCCGCGTCGCCACCTC 68039

Qy        629 CCAGGCCCGGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGC 688

Db      68038 CGTCGCCTGGGGCGCCTGGGACGGCGGCGGCATGTCCAGCGCCACCCCGGACATCGCCGC 67979

Qy        689 CCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCT 748

Db      67978 CCAGCTCGCCCAGGCTGGCCTGCGCCTGATGCCCCCGGACCTGGCGGTGTCCGCGCTGCT 67919

Qy        749 GTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCG 808

Db      67918 CGACGCCGTCGCCCGCGACGAGAGCACCACCGTCGTGACCGACGTGCGCTGGGAGGCGTT 67859

Qy        809 CGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGAC 868

Db      67858 CGCCGCCCGGTTCACCGCCCTGCGCCGCAGCCCCCTGCTCGCCGACCTGCCCGAGGCCCG 67799

Qy        869 GCCGGGCGACATCGACCCGATC 890

Db      67798 CCCCGAACCGGTCGCCGCCACC 67777


RESULT 4
US-11-052-554A-541/c
; Sequence 541, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
;  APPLICANT: Sachdeva, et al.
;  TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
;  TITLE OF INVENTION:  PROTEINS OF THERAPEUTIC POTENTIAL
;  FILE REFERENCE: 30853/40359A
;  CURRENT APPLICATION NUMBER: US/11/052,554A
;  CURRENT FILING DATE:  2005-02-07
;  PRIOR APPLICATION NUMBER: US 60/589,227
;  PRIOR FILING DATE: 2004-07-20
;  PRIOR APPLICATION NUMBER: IN 173/DEL/2004
;  PRIOR FILING DATE: 2004-02-06
;  NUMBER OF SEQ ID NOS: 763
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 541
;   LENGTH: 2352
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-541

  Query Match             7.9%;  Score 79.2;  DB 14;  Length 2352;
  Best Local Similarity   47.6%;  Pred. No. 9.6e-07;
  Matches  234;  Conservative    0;  Mismatches  258;  Indels    0;  Gaps    0;

Qy         50 CGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCG 109

Db       2008 CGCCGGCCCCGCCGTCCCCACCGGTGTCCAGGTTTTCGCCGTCACCGCCGGCTCCGCCGG 1949

Qy         110 GTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCAC 169

Db       1948 CGCCGCCCGGCCCGGCGAGCATTCCGGCATTCCCGCCGGCCCCGCCAACCCCGCCGTCGC 1889

Qy         170 CTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCA 229

Db       1888 CATTAGTGCTGAGCCCGCCGGTCCCGCCGTGACCTCCGCCGGCACCGATCAGCCCGGCCA 1829

Qy         230 CCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTC 289

Db       1828 GCAGTCCGCCGGTCCCGCCGGCCCCACCGGCCCCGCCGTCGCCACCAACAAGGCCGGTTC 1769

Qy         290 CGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGC 349

Db       1768 CACCGGCCCCGCCGGCCCCACCGGCGCCCCACAGCCCACCGACCCCGCCTCCCCCGCCGG 1709

Qy         350 CCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCA 409

Db       1708 TGCCGCCGAGTACGGTGCTGGCTCCGCCGATCCCGCCGGCCCCGCCGTCGCCGAGCAGCC 1649

Qy         410 CGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGC 469

Db       1648 AGCCGCCGGCCCCGCCGGCCCCACCGGCCCCGCCGCCACCCGCCGAGCTGCCTCCGGCCC 1589

Qy         470 CCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTC 529

Db       1588 CGCCGGCCCCGCCGGTACCCCACAGCCCGGCAGCCCCGCCCGCCCCGCCGGGCGCGCCCG 1529

Qy         530 ACGCGGGAGCTC 541

Db       1528 AGCCCGCCGCGC 1517


RESULT 5
US-10-501-035-144
; Sequence 144, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
;  TITLE OF INVENTION:  ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASES
;  TITLE OF INVENTION:  AND/OR PROTEIN TYROSINE KINASE PATHWAYS
;  FILE REFERENCE: D0185 PCT
;  CURRENT APPLICATION NUMBER: US/10/501,035
;  CURRENT FILING DATE:  2004-07-09
;  PRIOR APPLICATION NUMBER: US 60/350,061
;  PRIOR FILING DATE: 2002-01-18
;  NUMBER OF SEQ ID NOS: 795
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 144
;   LENGTH: 1738
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-501-035-144

  Query Match             7.9%;  Score 78.8;  DB 9;  Length 1738;
  Best Local Similarity   51.1%;  Pred. No. 1.2e-06;
  Matches  185;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy         149 CGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAG 208

Db         280 CGCCCAGAACCCCCTGCAGGTCGGGGCTGAGCTCCAGTCCCGCTTCTTTGCCTCCCAGGG 339

Qy         209 CAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGAT 268

Db         340 CTGCGCCCAGAGTCCATTCCAGGCCGCGCCGGCGCCCCCGCCCACGCCCCAGGCCCCGGC 399

Qy         269 CGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGC 328

Db         400 GGCCGAGCCCCTCCAGGTGGACTTGCTCCCGGTGCTCGCCGCCGCCCAGGAGTCCGCCGC 459
```

```
Qy        453 GGGGCGGACAGCATCGCCCCGCAGCCCGAGCCG------GGGCCCAACCCCGACGAGCTC 506
Db      65027 GTCTCGGAGCTGATGGGCTCGCTGAGCATGTCGTCGATCCGGGACTCCCTCGGCACGCTC 65086

Qy        507 GTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGG 566
Db      65087 CTGCCGATCCGGCGCACCGGCGAGCTGCCGCCGCTGTTCTGCGTCCATCCGGCCGGCGGG 65146

Qy        567 CTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCC 626
Db      65147 CTCAGCTGGTGCTACCTGCCGCTGGCCCGGCACGTGCCGGCCGACCGCCCGATCTACGGT 65206

Qy        627 TTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCG 686
Db      65207 CTGCAGGCGCGCGGCGCCGACGGCCGGGAGCCGCTCGCACCGTCGCTGCGCGAGATGGCC 65266

Qy        687 GCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGG 746
Db      65267 GCCGACTACGTGAGCCGGATGCGCGCCGTGCAGCCCGAGGGGCCGTACCACGTGCTCGGC 65326

Qy        747 CTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAG 806
Db      65327 TTCTCCTTCGGCGTGGCGCCCGCGCACGAGATCGCCGTCCAGCTGCGCGAGCAGGGCGCC 65386

Qy        807 CGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAG 866
Db      65387 GAGGTCGTGCTGGTGCTCATGGACTCCTATCCCATGGAGGATGCGGAGTCCGGCGAGCAG 65446

Qy        867 ACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCC 917
Db      65447 GCGGCCGACGAGGAGGAGCTGCCGTGGGAGGAGCTCATCGAGGCCGAGTTC 65497
```

RESULT 2
US-10-496-351-3
; Sequence 3, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4770
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-3

Query Match 8.0%; Score 79.6; DB 7; Length 4770;
Best Local Similarity 44.2%; Pred. No. 6.8e-07;
Matches 328; Conservative 0; Mismatches 414; Indels 0; Gaps 0;

```
Qy        149 CGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAG 208
Db       3462 CGCCGCGCACGTGGCCCGCCGCCTCGCCGCCCTCGGCGCCGACCACCTCGTCCTCACCAG 3521

Qy        209 CAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGAT 268
Db       3522 CCGCCGCGGCCCCTCGGCCCCCGGCGCGCCCGCGCTCGCCGCCGACCTGGAGGCGCTGGG 3581

Qy        269 CGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGC 328
Db       3582 CGCCCGCGTCACCACCGCCGCCTGCGACGTCGCCGACCGCGCCGCGCTCGCCGCCCTGCT 3641

Qy        329 AGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTC 388
Db       3642 GACCGGCCTGGAGCGCGCGGGCGAGCAGGTCACCGCCGTCGTGCACGCGGCGGGCGCCAA 3701

Qy        389 CGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACG 448
Db       3702 CGCCCAGACCCCCGTCGCCGACACCACCCCCGAGGAGCACGCGCGCGTCCAGGCCGCCAA 3761

Qy        449 GCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGT 508
Db       3762 GGCGCTCGGCGCCGAGCACCTCGACGAGCTGCTGGGCGGGCGCCCGCTCGACGCGTTCGT 3821

Qy        509 CGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCT 568
Db       3822 CCTGTTCTCCTCCAACGCGGGCGTGTGGGGCAGCGGCGGCCAGTGCGCCTACGCCGCCGC 3881

Qy        569 CGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTT 628
Db       3882 CAACGCCCGCCTCGACGCACTGGCCCAGCGCCGCCGCGCGCACGGCCGCGTCGCCACCTC 3941

Qy        629 CCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGC 688
Db       3942 CGTCGCCTGGGGCGCCTGGGACGGCGGCGGCATGTCCAGCGCCACCCCGGACATCGCCGC 4001

Qy        689 CCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCT 748
Db       4002 CCAGCTCGCCCAGGCTGGCCTGCGCCTGATGCCCCCGGACCTGGCGGTGTCCGCGCTGCT 4061

Qy        749 GTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCG 808
Db       4062 CGACGCCGTCGCCCGCGACGAGAGCACCACCGTCGTGACCGACGTGCGCTGGGAGGCGTT 4121

Qy        809 CGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGAC 868
Db       4122 CGCCGCCCCGGTTCACCGCCCTGCGCCGCAGCCCCCTGCTCGCCGACCTGCCCGAGGCCCG 4181

Qy        869 GCCGGGCGACATCGACCCGATC 890
Db       4182 CCCCGAACCGGTCGCCGCCACC 4203
```

RESULT 3
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 82746
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

Query Match 8.0%; Score 79.6; DB 7; Length 82746;
Best Local Similarity 44.2%; Pred. No. 3.5e-07;
Matches 328; Conservative 0; Mismatches 414; Indels 0; Gaps 0;

```
Qy        149 CGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAG 208
Db      68518 CGCCGCGCACGTGGCCCGCCGCCTCGCCGCCCTCGGCGCCGACCACCTCGTCCTCACCAG 68459

Qy        209 CAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGAT 268
Db      68458 CCGCCGCGGCCCCTCGGCCCCCGGCGCGCCCGCGCTCGCCGCCGACCTGGAGGCGCTGGG 68399

Qy        269 CGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGC 328
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 482.212 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_35000_35999
Perfect score: 1000
Sequence: 1 tggactgggagccgctcgac.....acccccgcccagcggctcta 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	76994	12	ADP64454	Adp64454 Sorangium
	2	173.6	17.4	7680	14	ACL72485	Acl72485 M. xanthu
c	3	173.6	17.4	11476	14	ACL64492	Acl64492 M. xanthu
	4	162.4	16.2	5463	14	ACL72687	Acl72687 M. xanthu
c	5	162.4	16.2	9710	14	ACL64515	Acl64515 M. xanthu
	6	137.6	13.8	25085	10	AAD54235	Aad54235 Streptomy
	7	137.6	13.8	50543	10	AAD54230	Aad54230 Streptomy
	8	129	12.9	22533	13	ADQ91713	Adq91713 Polyketid
	9	129	12.9	22533	13	AEB87005	Aeb87005 Streptomy
	10	129	12.9	22533	13	AEB86802	Aeb86802 Streptomy
	11	129	12.9	164051	13	ADQ91695	Adq91695 Polyketid
	12	129	12.9	164051	13	AEB86862	Aeb86862 Streptomy
	13	129	12.9	164051	13	AEB86659	Aeb86659 Streptomy
	14	125.8	12.6	24081	10	AAD54223	Aad54223 Streptomy
	15	125.8	12.6	52101	10	AAD54217	Aad54217 Streptomy
	16	124.8	12.5	125401	4	AAD17186	Aad17186 Streptomy
c	17	120.6	12.1	132544	13	ADV99887	Adv99887 Nanchangm
	18	120	12.0	44377	2	AAT80414	Aat80414 Platenoli
	19	120	12.0	44377	2	AAT78508	Aat78508 Platenoli
	20	115.2	11.5	65140	4	AAD17184	Aad17184 Streptomy
	21	114.6	11.5	5766	14	ADY72599	Ady72599 A. orient
	22	114.6	11.5	23496	14	ACL69182	Acl69182 M. xanthu
c	23	114.6	11.5	51354	14	ACL64807	Acl64807 M. xanthu
	24	114.6	11.5	73599	14	ADY72575	Ady72575 Polyene p
	25	110.6	11.1	113193	8	AAD54645	Aad54645 Streptomy
c	26	109.2	10.9	138203	14	ADZ51725	Adz51725 FR-008 po
	27	107	10.7	74787	12	ADQ74672	Adq74672 Streptomy
	28	105.2	10.5	15624	13	ADQ91707	Adq91707 Polyketid
	29	105.2	10.5	15624	13	AEB86999	Aeb86999 Streptomy
	30	105.2	10.5	15624	13	AEB86796	Aeb86796 Streptomy
	31	105	10.5	16299	13	ADQ91709	Adq91709 Polyketid
	32	105	10.5	16299	13	AEB87001	Aeb87001 Streptomy
	33	105	10.5	16299	13	AEB86798	Aeb86798 Streptomy
	34	103.8	10.4	64492	12	ADK16023	Adk16023 Streptomy
	35	103	10.3	4893	14	ACL71949	Acl71949 M. xanthu
	36	103	10.3	15482	14	ACL64604	Acl64604 M. xanthu
	37	102.2	10.2	77536	3	AAA14651	Aaa14651 Nucleotid
	38	100.2	10.0	84428	12	ADM45913	Adm45913 Streptomy
	39	96.2	9.6	53789	2	AAV21187	Aav21187 Amycolato
	40	95.8	9.6	9005	14	ACL71431	Acl71431 M. xanthu
	41	95.8	9.6	10793	14	ACL64599	Acl64599 M. xanthu
	42	95.6	9.6	10287	13	ADQ91701	Adq91701 Polyketid
	43	95.6	9.6	10287	13	AEB86993	Aeb86993 Streptomy
	44	95.6	9.6	10287	13	AEB86790	Aeb86790 Streptomy
	45	95	9.5	29467	14	ADY80051	Ady80051 Contig #1

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

PS Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC The present invention describes an isolated recombinant polynucleotide
CC (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC synthase (PKS) protein or a fragment comprising one or more domains of
CC the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC host cell (III) comprising (II), or (I) integrated into the cell
CC chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC the native disorazole PKS by the inactivation of one or more domain of
CC the disorazole PKS and/or addition of one or more domain of a non-
CC disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC expression system (VI) capable of producing a disorazole synthase domain
CC in a host cell, where the system comprises an encoding sequence for a
CC disorazole polyketide synthase domain, and the encoding sequence is
CC operably linked to control sequences effective in the cell to produce RNA
CC that is translated into the domain; (7) host cell modified to contain
CC (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC eliminate production of disorazole; (9) an isolated polypeptide encoded
CC by (I); (10) hybrid polyketide synthase comprising one or more
CC polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC of at least 200 base pairs with a sequence identical or substantially
CC identical to a protein encoding region of the 77294 base pair sequence of
CC SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC (I) is useful for producing disorazole polyketides. (III) is useful for
CC producing a polyketide, which involves growing (III) under conditions,
CC where a polyketide synthesised by a PKS comprising a protein encoded by
CC the polynucleotide molecule is produced in the cell. (I) is useful for
CC producing libraries of PKSs, where the polyketide producing colonies are
CC identified and isolated, and the produced polyketides are used
CC collectively in a panel to represent a library or assessed individually
CC for activity. The disorazole polyketide produced by (I) is useful in
CC molecular biology, chemistry, recombinant DNA technology, human and
CC veterinary medicine and agricultural applications. The disorazole
CC polyketide produced by (I) is formulated as a pharmaceutical composition,
CC useful for its activity against microorganisms, in a mammal. The present
CC sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

Query Match 100.0%; Score 1000; DB 12; Length 76994;
Best Local Similarity 100.0%; Pred. No. 2.5e-145;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35000 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 35059

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35060 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 35119

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35120 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 35179

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35180 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 35239

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35240 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 35299

Qy        301 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35300 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCG 35359

Qy        361 CCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35360 CCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCG 35419

Qy        421 TGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35420 TGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG 35479

Qy        481 AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35480 AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTT 35539

Qy        541 GCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35540 GCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACC 35599

Qy        601 ACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35600 ACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGG 35659

Qy        661 AGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35660 AGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCC 35719

Qy        721 TGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35720 TGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCC 35779

Qy        781 CCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35780 CCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCT 35839

Qy        841 TCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35840 TCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCA 35899

Qy        901 CGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35900 CGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCG 35959

Qy        961 CGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 1000
              ||||||||||||||||||||||||||||||||||||||||
Db      35960 CGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 35999
```

RESULT 2
ACL72485
ID ACL72485 standard; DNA; 7680 BP.
XX
AC ACL72485;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus gene sequence, seq id 8948.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression;
KW gene; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.

```
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 8948; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 7680 BP; 1248 A; 2466 C; 2790 G; 1176 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 173.6;  DB 14;  Length 7680;
  Best Local Similarity   52.5%;  Pred. No. 3e-18;
  Matches  532;  Conservative    0;  Mismatches  459;  Indels   23;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||| |  |||  | |   ||  |  ||  |||||        || ||||| || |
Db       2930 TCGACTGGGCGAAGCTGCATGGCGGCGCGCAGCTCCGGCGACGGAGCCTTCCCACCTATC 2989

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||||    |||   ||||||  |  | | |    |     |||     | || || || ||
Db       2990 CCTTCGCGAAGGAGAAGTACTGGGTAGCGAAGCCGGCGCGCGCGGAGAACCGCGTCCGGG 3049

Qy        121 CGGGAGAGAGGCG---CTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGC 177
              | ||     ||     | |||| ||||| ||||| | |||| | |||     | |   | ||
Db       3050 CCGGCAGCCGGGCGGCCCTGCATCCGCTGGTGCATGAGAACTCGTCCGACCTGCGGGGGC 3109

Qy        178 ACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACG 237
              |    |||  |  ||  |||  ||| || | ||||| ||||   || ||  |||||  |
Db       3110 AGCGCTACAGCGCCCACCTCACGGGAGAGGAGTTCTTCCTCCGAGACCATGTCGTGGGCC 3169

Qy        238 GCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGG 297
              | ||| | || ||||  | ||| ||||||||||||||| |||||||| | || ||  |
Db       3170 AGGACAAGGTGCTGCCCGGCGTGGCCTTCCTGGAGATGGCGCGCGCCGCGGTGGTGAACG 3229

Qy        298 CGTCCGGGCGGCCGGTCCGCGACA---------TCCTCGACGTCGTCTGGACCGCGCCCG 348
              |    || | | ||   | || ||         | |   |||||||||| |   ||||
Db       3230 CCAGTGGCCTGACGCCTCCCGGCAAGGTGACACTGCGGAACGTCGTCTGGTCACGGCCCA 3289

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGG---CCGCCGGCG 405
              || |||||  || ||  |||    || |  | ||    ||   | |||     | ||||
Db       3290 TCATCGTGACCGACGCCCCGGTGCAGGTGCACATCGGTCTGGAGGCGGAGCAGGACGGCA 3349

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              | ||| |||||   | | |   | | ||||    |||  | |   |  |   |||||||
Db       3350 GCGTCGTCTTCGAAATCTACGGAGGCACCGACGAGGCGCCGCTGCTCTTCTGTCAGGGGC 3409

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
              ||  |||| ||| ||  |||     ||  || |   | || ||   | |    ||     |
Db       3410 GGGCGCGCATCGGCGCTGGGACGGACGAGGAGGCCACTGCACCTTGCGTGGACCTGTCGG 3469

Qy        526 ACATCCTCTCACGTTGCTCG---AGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCC 582
              | |||  ||  ||||   |    ||       |||| |  ||   |||||      ||
Db       3470 TCCTCCAGTCCGGTTGTCAGGACCGGCGACAGAGCGGCGCGGCCGTCTACCAGACGTTCG 3529

Qy        583 AGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCG 642
              | |  |||||||  | ||||   ||   | || ||||||| |||||       ||
Db       3530 CGTCGGCCGGGCTGACCTACGGACCCACGCTTCGTGCGCTCCAAGAGCTCTCGGTGGGTC 3589

Qy        643 AGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCT 702
              |||  |||   | |||  |      |||||  ||||                |       |
Db       3590 GTGGCCAGGTGCTGGCGCGGGCCGTGCTCCCCGAGCAGGTGTTGAGCAGCCAGGGGCAGT 3649

Qy        703 TTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACG 762
              | | ||| |||||   | |||| || || ||  ||||     | ||||| ||||     |
Db       3650 TCGTCCTGCATCCGAGCTTGCTGGATGCCGCGTTGCAGGCGGCCCTCGGTCTGCACTCGG 3709

Qy        763 ATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCC 822
              | ||||  |  |   | |  ||          ||  |  | ||    |  | | | | |
Db       3710 ACGAGGGAGGACAGCTCTCGCTTCCGTTCGCGCTCGATGCGCTGGAGGTGTGGGACGGGC 3769

Qy        823 CGCTGCCGGCGCGGCTC----TTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACG 878
              |    ||  ||    ||   ||| |||  || | |     | | |||||| ||  |
Db       3770 CTGCTCCCTCGGACGTCTGGATTCACCTGCGCGATGAGACGGACGGCGCCGCGGGTAAGA 3829

Qy        879 AACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATG 938
              | || |||    | ||| | || | ||    || | ||| |||| |||     ||    ||
Db       3830 ACCGGCAGAAGATTTC-CATCTCCCTCTGTGACCGCGACGGCAAGGTCTGGGTCCGCCTG 3888

Qy        939 CGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992
              |  ||| ||   || | || | | |||| ||  | || | |   |||| ||
Db       3889 AAGGGATTCTCCACGCGGCCCGTGCCCGCGGCCGGTGTGCGCGCTGGCGCCGAG 3942


RESULT 3
ACL64492/c
ID   ACL64492 standard; DNA; 11476 BP.
XX
AC   ACL64492;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus DNA fragment, seq id 955.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 1; SEQ ID NO 955; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC   set of 1849 contig and singleton sequences comprising coding sequences,
CC   DNA replication elements, promoters and other regulatory elements from
CC   the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO
XX
SQ   Sequence 11476 BP; 1775 A; 4105 C; 3739 G; 1857 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 173.6;  DB 14;  Length 11476;
```

```
  Best Local Similarity   52.5%;  Pred. No. 2.8e-18;
  Matches  532;  Conservative    0;  Mismatches  459;  Indels   23;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | |||||||  |  |||  | |   ||  |  ||  |||||        || ||||| || |
Db       6502 TCGACTGGGCGAAGCTGCATGGCGGCGCGCAGCTCCGGCGACGGAGCCTTCCCACCTATC 6443

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||||    |||   |||||| |  | | |    |    |||    | || || || ||
Db       6442 CCTTCGCGAAGGAGAAGTACTGGGTAGCGAAGCCGGCGCGCGCGGAGAACCGCGTCCGGG 6383

Qy        121 CGGGAGAGAGGCG---CTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGC 177
              | ||      ||      | |||| ||||| ||||| | |||| | |||    | |  | ||
Db       6382 CCGGCAGCCGGGCGGCCCTGCATCCGCTGGTGCATGAGAACTCGTCCGACCTGCGGGGGC 6323

Qy        178 ACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACG 237
              |    |||  |  ||  |||  ||| || | ||||| ||||   || ||  |||||  |
Db       6322 AGCGCTACAGCGCCCACCTCACGGGAGAGGAGTTCTTCCTCCGAGACCATGTCGTGGGCC 6263

Qy        238 GCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGG 297
              | ||| | || |||| | ||| ||||||||||||||| |||||||| | || ||  |
Db       6262 AGGACAAGGTGCTGCCCGGCGTGGCCTTCCTGGAGATGGCGCGCGCCGCGGTGGTGAACG 6203

Qy        298 CGTCCGGGCGGCCGGTCCGCGACA---------TCCTCGACGTCGTCTGGACCGCGCCCG 348
              |    || | | ||   | || ||         | |   |||||||||| |   ||||
Db       6202 CCAGTGGCCTGACGCCTCCCGGCAAGGTGACACTGCGGAACGTCGTCTGGTCACGGCCCA 6143

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGG---CCGCCGGCG 405
              || |||||  || ||  |||    || |  | ||    ||   | |||     | ||||
Db       6142 TCATCGTGACCGACGCCCCGGTGCAGGTGCACATCGGTCTGGAGGCGGAGCAGGACGGCA 6083

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              | ||| |||||   | | |   | | ||||    |||  | |   |  |   |||||||
Db       6082 GCGTCGTCTTCGAAATCTACGGAGGCACCGACGAGGCGCCGCTGCTCTTCTGTCAGGGGC 6023

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
              ||  |||| ||| ||  |||      ||  || |   | || ||   | |    ||    |
Db       6022 GGGCGCGCATCGGCGCTGGGACGGACGAGGAGGCCACTGCACCTTGCGTGGACCTGTCGG 5963

Qy        526 ACATCCTCTCACGTTGCTCG---AGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCC 582
              | |||  ||  ||||   |   ||       |||| |  ||   |||||      ||
Db       5962 TCCTCCAGTCCGGTTGTCAGGACCGGCGACAGAGCGGCGCGGCCGTCTACCAGACGTTCG 5903

Qy        583 AGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCG 642
              |  |  |||||||  | ||||   ||   | || ||||||| ||||        ||
Db       5902 CGTCGGCCGGGCTGACCTACGGACCCACGCTTCGTGCGCTCCAAGAGCTCTCGGTGGGTC 5843

Qy        643 AGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCT 702
              |||  |||   | |||  |      ||||| ||||              |        |
Db       5842 GTGGCCAGGTGCTGGCGCGGGCCGTGCTCCCCGAGCAGGTGTTGAGCAGCCAGGGGCAGT 5783

Qy        703 TTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACG 762
              | | ||| |||||   | |||| || || ||  ||||    | ||||| ||||      |
Db       5782 TCGTCCTGCATCCGAGCTTGCTGGATGCCGCGTTGCAGGCGGCCCTCGGTCTGCACTCGG 5723

Qy        763 ATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCC 822
              | ||||  |  |    | |  ||          ||  |  | ||   |  |  | | | |
Db       5722 ACGAGGGAGGACAGCTCTCGCTTCCGTTCGCGCTCGATGCGCTGGAGGTGTGGGACGGGC 5663

Qy        823 CGCTGCCGGCGCGGCTC----TTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACG 878
              |    ||  ||    ||    ||| ||| || | |     | | ||||||| ||  |
Db       5662 CTGCTCCCTCGGACGTCTGGATTCACCTGCGCGATGAGACGGACGGCGCCGCGGGTAAGA 5603

Qy        879 AACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATG 938
              | || |||    | || | || | ||    || | || |||| |||     ||    ||
Db       5602 ACCGGCAGAAGATTTC-CATCTCCCTCTGTGACCGCGACGGCAAGGTCTGGGTCCGCCTG 5544

Qy        939 CGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992
              |  ||| ||   || | || | |||| ||  | || | |   |||| ||
Db       5543 AAGGGATTCTCCACGCGGCCCGTGCCCGCGGCCGGTGTGCGCGCTGGCGCCGAG 5490

RESULT 4

ACL72687
ID   ACL72687 standard; DNA; 5463 BP.
XX
AC   ACL72687;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus gene sequence, seq id 9150.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression;
KW   gene; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 9150; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 5463 BP; 807 A; 1753 C; 1987 G; 916 T; 0 U; 0 Other;

  Query Match            16.2%;  Score 162.4;  DB 14;  Length 5463;
  Best Local Similarity   51.7%;  Pred. No. 1.6e-16;
  Matches  432;  Conservative    0;  Mismatches  386;  Indels   18;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||  || |||  |||||    |   |||  || ||  | |  || ||   |||||
Db       2909 TCGACTGGCGGCGGCTGTACGCGACGCACACGCCTCGTCGGCTGCGCCTCCCGGGGTACC 2968

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              | ||| |     |||||||||       ||  |   |  |     |  | |  |    |
Db       2969 CGTTCTCGAAGAAGCGTTACTCCCCGCCCGTGGTGGAGGCGCAGGCCGTCTCCACGGTGA 3028

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              || |    |    | ||||| ||||| ||    |||| |||     |    | |||
Db       3029 GAGGCGGCGCGGTGCTTCACCCCCTCGTCCACCGCAACAGCTCCGTCCTGGCTGAGCAGC 3088

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              |||  |||| |||||||||||||| |||||||| || || || |  ||  |||| |
Db       3089 GCTACAGCAGCACGCTCCGGGGCGACGCCTTCTACCTGGCGGACCATCAGGTCAACGGGC 3148

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGC-----------GCCGCCG 288
              ||  |||| ||||| ||||| |  |||||||||| |||             | |||
Db       3149 GCAAAGTCCTGCCCGCCGCGGCCTATCTGGAGATGGCTCGCTTCGCGGCCAGTGACGCAT 3208
```

```
Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
Db       3209 GGGCGGACGAAGCGGTGGGCCGGATACCTGTCCTCAGGAACATGGTCTGGACCTCGCCCT 3268

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
Db       3269 TGGTGGTGGATGACGCGGCGGAGGTCGGAGTCGCGCTTCATGCGCAGGATGCGGGGCGGC 3328

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468
Db       3329 TGACGTTCGAGATTCATTCCGCGCCCGACGCAGCGGGGGAGCGCACCGTCCATTGCCGCG 3388

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528
Db       3389 GTCTGGCGGAGTTGCAGGAGCGCGTGGCGCCGCGCCATGACCTCCGGCGGCTGGAGGCGG 3448

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588
Db       3449 ATTGCACCCTGCGC------GTCCTGAGCCGGGAGCAGTGTTACGAGACCTTCGCCCGGC 3502

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648
Db       3503 TTGGTTTGAACTATGGGCCCGGTCACCGTGCGCTTGTCCGGGTGCGGGTGGGGCACGGCC 3562

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708
Db       3563 AAGCGCTGGGCGAGGTCGAGCTGCCCGAGCACCTGGTGGGCGACGGCCGATCGTTCGTGC 3622

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGG 768
Db       3623 TGCACCCCGCGCTGGTCGACTCCGCGCTGCAAGCGGTGGTCGGTGTGTTGGGCGACGAAG 3682

Qy        769 CGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCG 824
Db       3683 GCGGCACCGCGCCCGCGCTGCCGTTCGCAGTGGAAGAGGTCGTGGTCCATGGCCCG 3738
```

RESULT 5
ACL64515/c
ID ACL64515 standard; DNA; 9710 BP.
XX
AC ACL64515;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus DNA fragment, seq id 978.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.
XX
PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 1; SEQ ID NO 978; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC set of 1849 contig and singleton sequences comprising coding sequences,
CC DNA replication elements, promoters and other regulatory elements from
CC the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO
XX
SQ Sequence 9710 BP; 1636 A; 3444 C; 3101 G; 1529 T; 0 U; 0 Other;

```
  Query Match             16.2%;  Score 162.4;  DB 14;  Length 9710;
  Best Local Similarity   51.7%;  Pred. No. 1.5e-16;
  Matches  432;  Conservative    0;  Mismatches  386;  Indels   18;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
Db       2556 TCGACTGGCGGCGGCTGTACGCGACGCACACGCCTCGTCGGCTGCGCCTCCCGGGGTACC 2497

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
Db       2496 CGTTCTCGAAGAAGCGTTACTCCCCGCCCGTGGTGGAGGCGCAGGCCGTCTCCACGGTGA 2437

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
Db       2436 GAGGCGGCGCGGTGCTTCACCCCCTCGTCCACCGCAACAGCTCCGTCCTGGCTGAGCAGC 2377

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
Db       2376 GCTACAGCAGCACGCTCCGGGGCGACGCCTTCTACCTGGCGGACCATCAGGTCAACGGGC 2317

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGC------------GCCGCCG 288
Db       2316 GCAAAGTCCTGCCCGCCGCGGCCTATCTGGAGATGGCTCGCTTCGCGGCCAGTGACGCAT 2257

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
Db       2256 GGGCGGACGAAGCGGTGGGCCGGATACCTGTCCTCAGGAACATGGTCTGGACCTCGCCCT 2197

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
Db       2196 TGGTGGTGGATGACGCGGCGGAGGTCGGAGTCGCGCTTCATGCGCAGGATGCGGGGCGGC 2137

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468
Db       2136 TGACGTTCGAGATTCATTCCGCGCCCGACGCAGCGGGGGAGCGCACCGTCCATTGCCGCG 2077

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528
Db       2076 GTCTGGCGGAGTTGCAGGAGCGCGTGGCGCCGCGCCATGACCTCCGGCGGCTGGAGGCGG 2017

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588
Db       2016 ATTGCACCCTGCGC------GTCCTGAGCCGGGAGCAGTGTTACGAGACCTTCGCCCGGC 1963

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648
Db       1962 TTGGTTTGAACTATGGGCCCGGTCACCGTGCGCTTGTCCGGGTGCGGGTGGGGCACGGCC 1903

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708
Db       1902 AAGCGCTGGGCGAGGTCGAGCTGCCCGAGCACCTGGTGGGCGACGGCCGATCGTTCGTGC 1843

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGG 768
Db       1842 TGCACCCCGCGCTGGTCGACTCCGCGCTGCAAGCGGTGGTCGGTGTGTTGGGCGACGAAG 1783

Qy        769 CGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCG 824
Db       1782 GCGGCACCGCGCCCGCGCTGCCGTTCGCAGTGGAAGAGGTCGTGGTCCATGGCCCG 1727
```

```
RESULT 6
AAD54235
ID   AAD54235 standard; DNA; 25085 BP.
XX
AC   AAD54235;
XX
DT   17-JUN-2003  (first entry)
XX
DE   Streptomyces amphibiosporus lactimidomycin ORF5 DNA.
XX
KW   Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
KW   ds.
XX
OS   Streptomyces amphibiosporus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .25085
FT                   /*tag= a
FT                   /product= "LACT ORF5 protein"
FT                   /transl_except= (pos:5989. .5996, aa:Tyr-Arg)
FT                   /note= "No start codon"
FT                   /partial
XX
PN   WO200288176-A2.
XX
PD   07-NOV-2002.
XX
PF   26-APR-2002; 2002WO-CA000591.
XX
PR   26-APR-2001; 2001US-0286346P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazopoulos E,  Staffa A,  Yang X;
XX
DR   WPI; 2003-201222/19.
DR   P-PSDB; AAE35499.
XX
PT   Novel isolated or purified polypeptide involved in biosynthesis of
PT   polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
PT   dorrigocin or lactimidomycin.
XX
PS   Claim 6; Page 280-294; 312pp; English.
XX
CC   The invention relates to novel proteins involved in the biosynthesis of
CC   polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC   microorganisms. Sequences of the invention allow direct manipulation of
CC   dorrigocin, lactimidomycin and related chemical structures via chemical
CC   engineering of the enzymes involved in the biosynthesis of dorrigocin and
CC   lactimidomycin. They are useful for introducing chemical handles into
CC   normally inert positions that permit subsequence chemical modifications
CC   and facilitate the development of polyketides. The genes and proteins of
CC   the invention can also be used to generate a focused library of analogues
CC   around a polyketide lead candidate to fine-tune the compound for optimal
CC   properties. They are useful for generating antibodies specific for the
CC   polyketide biosynthesis. The present sequence is S. amphibiosporus
CC   lactimidomycin ORF5 DNA
XX
SQ   Sequence 25085 BP; 3179 A; 9400 C; 9226 G; 3280 T; 0 U; 0 Other;

  Query Match             13.8%;  Score 137.6;  DB 10;  Length 25085;
  Best Local Similarity   50.1%;  Pred. No. 8.4e-13;
  Matches  514;  Conservative    0;  Mismatches  464;  Indels   48;  Gaps    5;

Qy          3 GACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCC 62
              ||||||  | ||||    |     |||   | ||  ||  || |  |||| |||||||||
Db      23886 GACTGGCGGACGCTGTGGGGCGAACGGCACCCGCAACGGATCTCCATGCCGACGTACCCC 23945

Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
              |||   | |  ||| ||||||  ||  || || |       ||    | | |||||  |
Db      23946 TTCCAACTCCGGCGCTACTGGCTCGACATGACCA------CCCCGGCGCACGGCCCGCAC 23999

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |      |  | |||| ||||| ||||   ||||||||||    |   || |||
Db      24000 GTCTCCCGCGGACTGCATCCGCTGGTGCACCGGAACACCTCCGACCTGAGCGAGCAGCGC 24059

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              ||| ||   |   ||   |||   | ||||||| ||||||| |||| ||||| ||| |
Db      24060 TACACCTCGCACTTCACCGGCCGGGAGTTCTACATCGCCGACCACCGGGTGCAGGGCGAA 24119

Qy        243 AAGCTCCTCCCCGCGGCCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
              || || ||||||  |||||||| |||||||||| |||||||||  |   | ||| |
Db      24120 CAGGTCGTCCCCGGCGCGGCGTTGCTGGAGATGGCACGCGCCGCCGCCGTCCTCGCGGCG 24179

Qy        303 GGGCGGCCGGTCCGCGAC------ATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTG 356
              ||  |  |||     ||||        | | | | || |||||| ||  |||  |    |
Db      24180 GGCGGTGCGGAGACCGACTGGGCGCTGCGCCAGGTGGTCTGGTCCCGGCCGCTGACGGCC 24239

Qy        357 GGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTC 416
              ||              || |  |         |  |  || ||| | |||||     | ||||
Db      24240 GGACGGCCCGTCGACGTGCACACCGCCGTGTCCGTGCGGGCGGACGGCGAACCGGCCTTC 24299

Qy        417 G---------CCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGG 467
              |         | | |  | ||| || ||| ||||  |||  ||||  ||  | |  ||
Db      24300 GAGATCTACACGGAGGGCCCCGGCGGCGAACGCGTCGTGCACTCCACCGGACGGCTGCAC 24359

Qy        468 ATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGAC 527
              |   || || |   |  |||        ||  |  ||     |||||   ||  | |
Db      24360 CGCAGGACCGCCGGGAACGCCGCCGAACTCCTGGACGGACCCGAACTCCCCGGCGGCGCC 24419

Qy        528 ATCCTCT-----------------------CACGTTGCTCGAGGGTCACCGGCGGAGAC 563
                | ||                       |   |   |  |||  || ||  ||
Db      24420 GGACACTTGGACGTCGCCGCGCTGCGTGCCCAGTGCGACGGCACCGTCCTCGACGCCGAG 24479

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623
              | |||||||  |||  ||        | || ||| |  ||||   ||   |||| ||  ||
Db      24480 GAGTGCTACGCCCGGTTCTCCGGCGTGGGCCTGGAGTACGGGCCCACGCTGCGAGCCGTC 24539

Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683
              |  |||| || | |||     |  ||||  | ||   | | || ||  |||
Db      24540 GAGACGCTGAGCGGCGGCACCCGGCAGGCGGTGGCCCGGCTGCGGCTGTCGGCCGCCGCG 24599

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743
              || |      ||   || ||||| || ||   |||||||||||| || || || | |
Db      24600 TCCGCGAGG---ACCGGGTTCGCCCTGCACCCGAGCCTGCTCGACGCCGCACTCCAGTCC 24656

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803
              |||  || ||  |  || |  |  |     |  |    || || || ||| | | |
Db      24657 ACGGCGGGCCTCTTCACCGGTTCCGGTACGTCCTCGGCGGCCCTGCCGTTTGCCCTGGAC 24716

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863
              ||| |   | |||| ||| ||  |||  |   |   |  ||     |  ||
Db      24717 CGGCTGGAGGTGCTGCGCGCGACGCCCTCCTCGGGGTGGGCGGTGGCACGCTTCGCCGCC 24776

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923
              |     |          |   |   |    || ||    ||  |||   || ||||  |
Db      24777 GACGACCGCCCAGGCGGGGTGCGCCGCCTGGACATCGACGTGTGCGACGACGACGGCGAG 24836

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACC 983
              |||  | |  ||| || |  |||   |||||          |  |||| ||| ||  ||
Db      24837 GTGTGCGTACGGATCCGGGGCTTCCAGGTCCGTACGTACGGCGGCGACGCCGCCCCGTCC 24896

Qy        984 CCCGCC 989
              || ||
Db      24897 GCCTCC 24902


RESULT 7
AAD54230
ID   AAD54230 standard; DNA; 50543 BP.
XX
AC   AAD54230;
XX
DT   17-JUN-2003  (first entry)
```

```
XX
DE   Streptomyces amphibiosporus lactimidomycin DNA.
XX
KW   Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
KW   ds.
XX
OS   Streptomyces amphibiosporus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1698
FT                   /*tag= a
FT                   /product= "LACT ORF1 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             1908. .2162
FT                   /*tag= b
FT                   /product= "LACT ORF2 protein"
FT   CDS             2166. .4136
FT                   /*tag= c
FT                   /product= "LACT ORF3 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             4152. .14462
FT                   /*tag= d
FT                   /product= "LACT ORF4 protein"
FT   CDS             14549. .39631
FT                   /*tag= e
FT                   /product= "LACT ORF5 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             39628. .45924
FT                   /*tag= f
FT                   /product= "LACT ORF6 protein"
FT   CDS             45926. .48232
FT                   /*tag= g
FT                   /product= "LACT ORF7 protein"
FT   CDS             48441. .49697
FT                   /*tag= h
FT                   /product= "LACT ORF8 protein"
FT   CDS             complement(49800. .40543)
FT                   /*tag= i
FT                   /product= "LACT ORF9 protein"
XX
PN   WO200288176-A2.
XX
PD   07-NOV-2002.
XX
PF   26-APR-2002; 2002WO-CA000591.
XX
PR   26-APR-2001; 2001US-0286346P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazopoulos E,  Staffa A,  Yang X;
XX
DR   WPI; 2003-201222/19.
DR   P-PSDB; AAE35496, AAE35501, AAE35502, AAE35503, AAE34595, AAE34597,
DR   AAE34598, AAE34599, AAE34500.
XX
PT   Novel isolated or purified polypeptide involved in biosynthesis of
PT   polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
PT   dorrigocin or lactimidomycin.
XX
PS   Claim 1; Page 201-229; 312pp; English.
XX
CC   The invention relates to novel proteins involved in the biosynthesis of
CC   polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC   microorganisms. Sequences of the invention allow direct manipulation of
CC   dorrigocin, lactimidomycin and related chemical structures via chemical
CC   engineering of the enzymes involved in the biosynthesis of dorrigocin and
CC   lactimidomycin. They are useful for introducing chemical handles into
CC   normally inert positions that permit subsequence chemical modifications
CC   and facilitate the development of polyketides. The genes and proteins of
CC   the invention can also be used to generate a focused library of analogues
CC   around a polyketide lead candidate to fine-tune the compound for optimal
CC   properties. They are useful for generating antibodies specific for the
CC   polyketide biosynthesis. The present sequence is S. amphibiosporus
CC   lactimidomycin DNA
XX
SQ   Sequence 50543 BP; 6619 A; 18841 C; 18399 G; 6684 T; 0 U; 0 Other;

  Query Match           13.8%;  Score 137.6;  DB 10;  Length 50543;
  Best Local Similarity   50.1%;  Pred. No. 7.6e-13;
  Matches  514;  Conservative    0;  Mismatches  464;  Indels   48;  Gaps    5;

Qy          3 GACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCC 62
              ||||||  | ||||    |     |||   | ||  ||  || |  |||| |||||||||
Db      38432 GACTGGCGGACGCTGTGGGGCGAACGGCACCCGCAACGGATCTCCATGCCGACGTACCCC 38491

Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
              |||   | |  ||| |||||| ||  || || |       ||    | | ||||| |
Db      38492 TTCCAACTCCGGCGCTACTGGCTCGACATGACCA------CCCCGGCGCACGGCCCGCAC 38545

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |       |  |  |||| ||||| |||||    ||||||||||    |   || |||
Db      38546 GTCTCCCGCGGACTGCATCCGCTGGTGCACCGGAACACCTCCGACCTGAGCGAGCAGCGC 38605

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              ||| ||    |   ||   |||   | |||||| ||||||| |||| ||||| ||| |
Db      38606 TACACCTCGCACTTCACCGGCCGGGAGTTCTACATCGCCGACCACCGGGTGCAGGGCGAA 38665

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGTCC 302
              || || ||||||  |||||||| |||||||||| ||||||||||  |    | ||| |
Db      38666 CAGGTCGTCCCCGGCGCGGCGTTGCTGGAGATGGCACGCGCCGCCGCCGTCCTCGCGGCG 38725

Qy        303 GGGCGGCCGGTCCGCGAC------ATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTG 356
              ||  |  |||    ||||       | | | | || |||||| ||  |||  |   |
Db      38726 GGCGGTGCGGAGACCGACTGGGCGCTGCGCCAGGTGGTCTGGTCCCGGCCGCTGACGGCC 38785

Qy        357 GGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTC 416
              ||              || | |          |  |  || ||| | |||||     | ||||
Db      38786 GGACGGCCCGTCGACGTGCACACCGCCGTGTCCGTGCGGGCGGACGGCGAACCGGCCTTC 38845

Qy        417 G---------CCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGG 467
              |         | | |  | ||| || ||| ||||  |||  ||||  ||  | |  ||
Db      38846 GAGATCTACACGGAGGGCCCCGGCGGCGAACGCGTCGTGCACTCCACCGGACGGCTGCAC 38905

Qy        468 ATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGAC 527
              |   || || |   |  |||       ||  |  ||    |||||   ||  | |
Db      38906 CGCAGGACCGCCGGGAACGCCGCCGAACTCCTGGACGGACCCGAACTCCCCGGCGGCGCC 38965

Qy        528 ATCCTCT----------------------CACGTTGCTCGAGGGTCACCGGCGGAGAC 563
                | ||                      |   |   | |||  || ||  ||
Db      38966 GGACACTTGGACGTCGCCGCGCTGCGTGCCCAGTGCGACGGCACCGTCCTCGACGCCGAG 39025

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623
              | |||||||  |||  ||        | || ||| |  ||||   ||   |||| ||  ||
Db      39026 GAGTGCTACGCCCGGTTCTCCGGCGTGGGCCTGGAGTACGGGCCCACGCTGCGAGCCGTC 39085

Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683
                |  |||| || | |||    |  ||||  | ||   | | || ||  |||
Db      39086 GAGACGCTGAGCGGCGGCACCCGGCAGGCGGTGGCCCGGCTGCGGCTGTCGGCCGCCGCG 39145

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743
              || |        ||   || ||||| || ||    |||||||||||| || || || | |
Db      39146 TCCGCGAGG---ACCGGGTTCGCCCTGCACCCGAGCCTGCTCGACGCCGCACTCCAGTCC 39202

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803
              |||   || ||  |   || |   |   |   |  |    || || || ||| | | |
Db      39203 ACGGCGGGCCTCTTCACCGGTTCCGGTACGTCCTCGGCGGCCCTGCCGTTTGCCCTGGAC 39262

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863
              ||| |   | |||| ||| ||  |||  |   |   |  ||    ||        |  ||
Db      39263 CGGCTGGAGGTGCTGCGCGCGACGCCCTCCTCGGGGTGGGCGGTGGCACGCTTCGCCGCC 39322

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923
```

```
Db      39323 GACGACCGCCCAGGCGGGGTGCGCCGCCTGGACATCGACGTGTGCGACGACGACGGCGAG 39382

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACC 983

Db      39383 GTGTGCGTACGGATCCGGGGCTTCCAGGTCCGTACGTACGGCGGCGACGCCGCCCCGTCC 39442

Qy        984 CCCGCC 989

Db      39443 GCCTCC 39448


RESULT 8
ADQ91713
ID   ADQ91713 standard; DNA; 22533 BP.
XX
AC   ADQ91713;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Polyketide synthase ORF17, SEQ ID 36.
XX
KW   Antifungal; Cytostatic; polyketide synthase; enzyme; polyene polyketide;
KW   polyketide; cancer; gene; ds.
XX
OS   Streptomyces aizunensis; NRRL B-11277.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .22533
FT                   /*tag=  a
FT                   /product= "Polyketide synthase ORF17 protein"
XX
PN   WO2004065401-A1.
XX
PD   05-AUG-2004.
XX
PF   21-JAN-2004; 2004WO-CA000068.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Bachmann BO,  Mcalpine JB,  Zazopoulos E,  Farnet CM;
XX
DR   WPI; 2004-348649/33.
DR   P-PSDB; ADQ91712.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Disclosure; SEQ ID NO 36; 554pp; English.
XX
CC   The present invention relates to novel Streptomyces aizunensis (NRRL B-
CC   11277) polyketide synthases and their coding sequences for polyene
CC   polyketide production. The polyketides are useful as fungal growth
CC   inhibitors and cancer cell growth inhibitors. The present sequence is a
CC   coding sequence for a polyketide synthase of the invention.
XX
SQ   Sequence 22533 BP; 2786 A; 8775 C; 7780 G; 3192 T; 0 U; 0 Other;

  Query Match             12.9%;  Score 129;  DB 13;  Length 22533;
  Best Local Similarity   49.9%;  Pred. No. 1.8e-11;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60

Db       8069 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 8128

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120

Db       8129 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 8188

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180

Db       8189 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 8248

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225

Db       8249 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 8308

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285

Db       8309 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 8368

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345

Db       8369 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 8428

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405

Db       8429 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 8488

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465

Db       8489 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 8545

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525

Db       8546 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 8605

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585

Db       8606 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 8665

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645

Db       8666 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 8725

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705

Db       8726 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 8785

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765

Db       8786 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 8845

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825

Db       8846 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 8903

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885

Db       8904 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 8963

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945

Db       8964 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 9022

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985

Db       9023 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 9063


RESULT 9
AEB87005
ID   AEB87005 standard; DNA; 22533 BP.
XX
AC   AEB87005;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Streptomyces aizunensis polyketide synthase ORF17 DNA SEQ ID NO:36.
XX
```

```
KW   polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW   gene; ds.
XX
OS   Streptomyces aizunensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .22533
FT                   /*tag=  a
FT                   /product= "polyene polyketide synthase ORF17"
XX
PN   CA2453071-A1.
XX
PD   03-APR-2004.
XX
PF   21-JAN-2004; 2004CA-02453071.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazapoulos E;
XX
DR   WPI; 2004-348649/33.
DR   P-PSDB; AEB87004.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Claim 2; SEQ ID NO 36; 550pp; English.
XX
CC   The invention relates to an isolated, purified or enriched nucleic acid
CC   which can be used for the production of a polyketide. The nucleic acid
CC   encodes at least one domain of the polyketide synthase system formed by
CC   the polyketide synthases of AEB86865, AEB86867, AEB86869, AEB86871,
CC   AEB86873, AEB86875, AEB86877, AEB86879 and AEB86881. Also described: (1)
CC   an isolated, purified or enriched nucleic acid for the production of a
CC   polyketide of Formula I comprising a nucleic acid selected from the group
CC   consisting of AEB86866, AEB86868, AEB86870, AEB86872, AEB86874, AEB86876,
CC   AEB86878, AEB86880 and AEB86882; (2) an isolated, purified or enriched
CC   nucleic acid for the production of a polyene polyketide; (3) an
CC   expression vector comprising the novel nucleic acid or the nucleic acid
CC   of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC   the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC   250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC   the production of a polyketide. The nucleic acid can be used for
CC   production of a polyketide, for use as a fungal and cancer cell growth
CC   inhibitors. The present sequence encodes Streptomyces aizunensis polyene
CC   polyketide synthase ORF17, which is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 22533 BP; 2786 A; 8775 C; 7780 G; 3192 T; 0 U; 0 Other;

  Query Match             12.9%;  Score 129;  DB 13;  Length 22533;
  Best Local Similarity   49.9%;  Pred. No. 1.8e-11;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||    ||  |  |||| ||   |     || ||||||   |||||||| |||
Db       8069 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 8128

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||   |||||||| | |||| ||          |  | |  | ||  |  || | |
Db       8129 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 8188

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              |||  |  | | ||  |  ||  | |  || |     | || ||     || |||  ||
Db       8189 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 8248

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225
                          || | || ||||||||  |  |||| |   |  |   |||||||| |

Db       8249 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 8308

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285
              ||  |||     ||| |     || || ||  |   | || ||| | ||  |  |||  |  ||
Db       8309 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 8368

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345
              | || ||   |||   | |    |   ||   ||  ||  |  ||  |    |     |
Db       8369 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 8428

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
              || |     || ||| |  |  | | |||||| | |       ||      |  |||||
Db       8429 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 8488

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              |  |  | | | ||||| ||||  || |  |  |||       | || ||| || | ||
Db       8489 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 8545

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
                | |  || ||          || |||  |    |||  ||  || ||   ||    |
Db       8546 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 8605

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
              | | | |   || |   | | ||||| ||||   ||||   | ||||  |
Db       8606 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 8665

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
              |  ||| |   | ||||  | | | |  ||| | ||    | |   ||  |
Db       8666 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 8725

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
              ||| |   ||| ||| |   ||| || ||    |  || |  ||   | ||   || |
Db       8726 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 8785

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              ||| || || || || |||||||| ||||||||   |||  |||  | ||| |  |  |
Db       8786 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 8845

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
              | | || ||     |   || || ||| ||    ||   ||  | |||| |   |||
Db       8846 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 8903

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885
              |  |    | || || ||| |  ||| | ||  || | ||   | |   | |    | |||
Db       8904 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 8963

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945
              |  | || ||  |||    ||       ||||| |   |  | |||   | ||||
Db       8964 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 9022

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985
              || ||| |  ||  | || | ||| |||   ||| |||| |
Db       9023 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 9063


RESULT 10
AEB86802
ID   AEB86802 standard; DNA; 22533 BP.
XX
AC   AEB86802;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Streptomyces aizunensis polyketide synthase ORF17 DNA SEQ ID NO:36.
XX
KW   polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW   gene; ds.
XX
OS   Streptomyces aizunensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .22533
FT                   /*tag=  a
```

```
FT                   /product= "polyene polyketide synthase ORF17"
XX
PN   CA2453080-A1.
XX
PD   03-APR-2004.
XX
PF   21-JAN-2004; 2004CA-02453080.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Bachmann BO,  Mcalpine JB,  Zazopoulos E;
XX
DR   WPI; 2004-348650/33.
DR   P-PSDB; AEB86801.
XX
PT   Polyene polyketide compounds are bacterial cell growth inhibitors useful
PT   for treating e.g. fungal infections and cancer.
XX
PS   Claim 2; SEQ ID NO 36; 144pp; English.
XX
CC   The invention relates to an isolated, purified or enriched nucleic acid
CC   which can be used for the production of a polyketide. The nucleic acid
CC   encodes at least one domain of the polyketide synthase system formed by
CC   the polyketide synthases of AEB86662, AEB86664, AEB86666, AEB86668,
CC   AEB86670, AEB86672, AEB86674, AEB86676 and AEB86678. Also described: (1)
CC   an isolated, purified or enriched nucleic acid for the production of a
CC   polyketide of Formula I comprising a nucleic acid selected from the group
CC   consisting of AEB86663, AEB86665, AEB86667, AEB86669, AEB86671, AEB86673,
CC   AEB86675, AEB86677 and AEB86679; (2) an isolated, purified or enriched
CC   nucleic acid for the production of a polyene polyketide; (3) an
CC   expression vector comprising the novel nucleic acid or the nucleic acid
CC   of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC   the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC   250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC   the production of a polyketide. The nucleic acid can be used for
CC   production of a polyketide, for use as a fungal and cancer cell growth
CC   inhibitors. The present sequence encodes Streptomyces aizunensis polyene
CC   polyketide synthase ORF17, which is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 22533 BP; 2786 A; 8775 C; 7780 G; 3192 T; 0 U; 0 Other;

  Query Match             12.9%;  Score 129;  DB 13;  Length 22533;
  Best Local Similarity   49.9%;  Pred. No. 1.8e-11;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||    ||  |  |||| ||   |      || ||||||   |||||||| |||
Db       8069 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 8128

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||   |||||||| | |||| ||          |  | |   | ||  |  || | |
Db       8129 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 8188

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              |||  |  | | ||  |  ||  | |  || |     | || ||     || |||  ||
Db       8189 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 8248

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225
                         || | || ||||||||  |  |||| |   |  |   |||||||| |
Db       8249 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 8308

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285
              ||  |||    ||| |    || || || |   | || ||| | ||  | |||  |  ||
Db       8309 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 8368

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345
              | || ||   |||   | |    |   ||   ||  ||  |  ||  |   |      |

Db       8369 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 8428

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
              || |     || ||| |  |  | | |||||| | |      ||      |  |||||
Db       8429 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 8488

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              |  |  | | | ||||| ||||  || |  |  |||      | || ||| || | ||
Db       8489 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 8545

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
              | |  || ||          || ||| |    ||| ||  || ||   ||     |
Db       8546 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 8605

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
              | | | |   || |    | |  |||||  ||||    ||||   | ||||   |
Db       8606 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 8665

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
              |  |||  |   | ||||   |  | | |  | ||| | ||     | |    ||  |
Db       8666 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 8725

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
              |||  |    ||| ||| |   ||| || ||     |  || |  ||   | ||   || |
Db       8726 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 8785

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              ||| || || || || |||||||| ||||||||   |||  |||  | || |  |   |
Db       8786 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 8845

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
              | | || ||     |   || || ||| ||    ||   ||  | |||| |   |||
Db       8846 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 8903

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885
              |  |   | || || ||| |  ||| | ||  || | ||   | |   | |   | |||
Db       8904 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 8963

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945
              |  | || ||  |||    ||      |||||  |   |  | |||   | ||||
Db       8964 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 9022

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985
              || ||| |  ||  | || | ||| |||   ||| |||| |
Db       9023 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 9063


RESULT 11
ADQ91695
ID   ADQ91695 standard; DNA; 164051 BP.
XX
AC   ADQ91695;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Polyketide synthase related DNA contig 2, SEQ ID 18.
XX
KW   Antifungal; Cytostatic; polyketide synthase; enzyme; polyene polyketide;
KW   polyketide; cancer; gene; ds.
XX
OS   Streptomyces aizunensis; NRRL B-11277.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(6. .1109)
FT                   /*tag=  a
FT                   /product= "Polyketide synthase related ORF9 protein"
FT   CDS             1375. .25818
FT                   /*tag=  b
FT                   /product= "Polyketide synthase ORF10 protein"
FT   CDS             25902. .36188
FT                   /*tag=  c
FT                   /product= "Polyketide synthase ORF11 protein"
FT   CDS             36213. .56468
```

```
FT                   /*tag=  d
FT                   /product= "Polyketide synthase ORF12 protein"
FT   CDS             56600. .61573
FT                   /*tag=  e
FT                   /product= "Polyketide synthase ORF13 protein"
FT   CDS             61852. .77475
FT                   /*tag=  f
FT                   /product= "Polyketide synthase ORF14 protein"
FT   CDS             77606. .93904
FT                   /*tag=  g
FT                   /product= "Polyketide synthase ORF15 protein"
FT   CDS             94057. .103740
FT                   /*tag=  h
FT                   /product= "Polyketide synthase ORF16 protein"
FT   CDS             103789. .126321
FT                   /*tag=  i
FT                   /product= "Polyketide synthase ORF17 protein"
FT   CDS             126389. .138007
FT                   /*tag=  j
FT                   /product= "Polyketide synthase ORF18 protein"
FT   CDS             complement(138063. .139079)
FT                   /*tag=  k
FT                   /product= "Polyketide synthase related ORF19 protein"
FT   CDS             complement(139266. .140117)
FT                   /*tag=  l
FT                   /product= "Polyketide synthase related ORF20 protein"
FT   CDS             complement(140114. .141103)
FT                   /*tag=  m
FT                   /product= "Polyketide synthase related ORF21 protein"
FT   CDS             141483. .142436
FT                   /*tag=  n
FT                   /product= "Polyketide synthase related ORF22 protein"
FT   CDS             142440. .143054
FT                   /*tag=  o
FT                   /product= "Polyketide synthase related ORF23 protein"
FT   CDS             143133. .144119
FT                   /*tag=  p
FT                   /product= "Polyketide synthase related ORF24 protein"
FT   CDS             144116. .145102
FT                   /*tag=  q
FT                   /product= "Polyketide synthase related ORF25 protein"
FT   CDS             145099. .145743
FT                   /*tag=  r
FT                   /product= "Polyketide synthase related ORF26 protein"
FT   CDS             145818. .147230
FT                   /*tag=  s
FT                   /product= "Polyketide synthase related ORF27 protein"
FT   CDS             complement(147306. .148967)
FT                   /*tag=  t
FT                   /product= "Polyketide synthase related ORF28 protein"
FT   CDS             complement(149176. .149871)
FT                   /*tag=  u
FT                   /product= "Polyketide synthase related ORF29 protein"
FT   CDS             complement(149868. .150788)
FT                   /*tag=  v
FT                   /product= "Polyketide synthase related ORF30 protein"
FT   CDS             complement(150769. .153765)
FT                   /*tag=  w
FT                   /product= "Polyketide synthase related ORF31 protein"
FT   CDS             154485. .156041
FT                   /*tag=  x
FT                   /product= "Polyketide synthase related ORF32 protein"
FT   CDS             156075. .157064
FT                   /*tag=  y
FT                   /product= "Polyketide synthase related ORF33 protein"
FT   CDS             157308. .158873
FT                   /*tag=  z
FT                   /product= "Polyketide synthase related ORF34 protein"
FT   CDS             158970. .160202
FT                   /*tag=  aa
FT                   /product= "Polyketide synthase related ORF35 protein"
FT   CDS             160199. .161719
FT                   /*tag=  ab
FT                   /product= "Polyketide synthase related ORF36 protein"
FT   CDS             161924. .162577
FT                   /*tag=  ac
FT                   /product= "Polyketide synthase related ORF37 protein"
FT   CDS             162723. .164051
FT                   /*tag=  ad
FT                   /product= "Polyketide synthase related ORF38 protein"
XX
PN   WO2004065401-A1.
XX
PD   05-AUG-2004.
XX
PF   21-JAN-2004; 2004WO-CA000068.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Bachmann BO,  Mcalpine JB,  Zazopoulos E,  Farnet CM;
XX
DR   WPI; 2004-348649/33.
DR   P-PSDB; ADQ91696, ADQ91698, ADQ91700, ADQ91702, ADQ91704, ADQ91706,
DR   ADQ91708, ADQ91710, ADQ91712, ADQ91714, ADQ91716, ADQ91718, ADQ91720,
DR   ADQ91722, ADQ91724, ADQ91726, ADQ91728, ADQ91730, ADQ91732, ADQ91734,
DR   ADQ91736, ADQ91738, ADQ91740, ADQ91742, ADQ91744, ADQ91746, ADQ91748,
DR   ADQ91750, ADQ91752, ADQ91754.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Disclosure; SEQ ID NO 18; 554pp; English.
XX
CC   The present invention relates to novel Streptomyces aizunensis (NRRL B-
CC   11277) polyketide synthases and their coding sequences for polyene
CC   polyketide production. The polyketides are useful as fungal growth
CC   inhibitors and cancer cell growth inhibitors. The present sequence is a
CC   DNA contig which comprises thirty Open Reading Frames (ORFs), encoding
CC   polyketide synthases and related proteins.
XX
SQ   Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;

  Query Match             12.9%;  Score 129;  DB 13;  Length 164051;
  Best Local Similarity   49.9%;  Pred. No. 1.3e-11;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||    ||  |  |||| ||    |      || ||||||   |||||||| |||
Db     111857 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 111916

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||   |||||||| | |||| ||         |  |  |  ||  |   ||  |
Db     111917 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 111976

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              |||  |  | | ||  |  ||  | |  || |     | || ||     || |||  ||
Db     111977 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 112036

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225
                          || | || ||||||||  |  |||| |   |  |   |||||||| |
Db     112037 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 112096

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285
              ||  |||    ||| |    || || || |   | || ||| | ||  | |||  |  ||
Db     112097 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 112156

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345
              | || ||   |||   | |    |   ||  ||  ||  |  ||  |    |       |
Db     112157 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 112216

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
```

```
Db     112217 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 112276

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465

Db     112277 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 112333

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525

Db     112334 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 112393

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585

Db     112394 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 112453

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645

Db     112454 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 112513

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705

Db     112514 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 112573

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765

Db     112574 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 112633

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825

Db     112634 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 112691

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885

Db     112692 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 112751

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945

Db     112752 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 112810

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985

Db     112811 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 112851


RESULT 12
AEB86862
ID   AEB86862 standard; DNA; 164051 BP.
XX
AC   AEB86862;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Streptomyces aizunensis polyketide synthase contig 2 SEQ ID NO:18.
XX
KW   polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW   ds.
XX
OS   Streptomyces aizunensis.
XX
PN   CA2453071-A1.
XX
PD   03-APR-2004.
XX
PF   21-JAN-2004; 2004CA-02453071.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazapoulos E;
XX
DR   WPI; 2004-348649/33.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Disclosure; SEQ ID NO 18; 550pp; English.
XX
CC   The invention relates to an isolated, purified or enriched nucleic acid
CC   which can be used for the production of a polyketide. The nucleic acid
CC   encodes at least one domain of the polyketide synthase system formed by
CC   the polyketide synthases of AEB86865, AEB86867, AEB86869, AEB86871,
CC   AEB86873, AEB86875, AEB86877, AEB86879 and AEB86881. Also described: (1)
CC   an isolated, purified or enriched nucleic acid for the production of a
CC   polyketide of Formula I comprising a nucleic acid selected from the group
CC   consisting of AEB86866, AEB86868, AEB86870, AEB86872, AEB86874, AEB86876,
CC   AEB86878, AEB86880 and AEB86882; (2) an isolated, purified or enriched
CC   nucleic acid for the production of a polyene polyketide; (3) an
CC   expression vector comprising the novel nucleic acid or the nucleic acid
CC   of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC   the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC   250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC   the production of a polyketide. The nucleic acid can be used for
CC   production of a polyketide, for use as a fungal and cancer cell growth
CC   inhibitors. The present sequence represents a Streptomyces aizunensis
CC   polyene polyketide synthase related contig 2 DNA sequence, which is used
CC   in the exemplification of the present invention.
XX
SQ   Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;

  Query Match             12.9%;  Score 129;  DB 13;  Length 164051;
  Best Local Similarity   49.9%;  Pred. No. 1.3e-11;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60

Db     111857 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 111916

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120

Db     111917 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 111976

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180

Db     111977 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 112036

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225

Db     112037 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 112096

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285

Db     112097 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 112156

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345

Db     112157 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 112216

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405

Db     112217 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 112276

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465

Db     112277 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 112333

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525

Db     112334 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 112393

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585

Db     112394 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 112453
```

```
Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
              |  |||  |   | ||||   |  | | |  | ||| | ||      | |    ||  |
Db     112454 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 112513

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
              ||| |    ||| ||| |   ||| || ||    |  || |  ||   | ||   || |
Db     112514 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 112573

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              ||| || || || || ||||||||| |||||||   |||  |||  | || |  |   |
Db     112574 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 112633

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
              | | || ||        |   || || ||| ||    ||   ||  | |||| |   |||
Db     112634 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 112691

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885
              |  |   | || || ||| |  ||| | ||  || | ||   | |   | |   | |||
Db     112692 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 112751

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945
              |  | || ||  |||    ||      ||||| |   |  | |||   | ||||
Db     112752 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 112810

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985
              || ||| |  ||  | || | ||| |||   ||| |||| |
Db     112811 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 112851
```

RESULT 13
AEB86659
ID AEB86659 standard; DNA; 164051 BP.
XX
AC AEB86659;
XX
DT 06-OCT-2005 (first entry)
XX
DE Streptomyces aizunensis polyketide synthase contig 2 SEQ ID NO:18.
XX
KW polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW ds.
XX
OS Streptomyces aizunensis.
XX
PN CA2453080-A1.
XX
PD 03-APR-2004.
XX
PF 21-JAN-2004; 2004CA-02453080.
XX
PR 21-JAN-2003; 2003US-0441123P.
PR 13-MAY-2003; 2003US-0469810P.
PR 01-AUG-2003; 2003US-0491516P.
PR 13-AUG-2003; 2003US-0494568P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Bachmann BO, Mcalpine JB, Zazopoulos E;
XX
DR WPI; 2004-348650/33.
XX
PT Polyene polyketide compounds are bacterial cell growth inhibitors useful
PT for treating e.g. fungal infections and cancer.
XX
PS Disclosure; SEQ ID NO 18; 144pp; English.
XX
CC The invention relates to an isolated, purified or enriched nucleic acid
CC which can be used for the production of a polyketide. The nucleic acid
CC encodes at least one domain of the polyketide synthase system formed by
CC the polyketide synthases of AEB86662, AEB86664, AEB86666, AEB86668,
CC AEB86670, AEB86672, AEB86674, AEB86676 and AEB86678. Also described: (1)
CC an isolated, purified or enriched nucleic acid for the production of a
CC polyketide of Formula I comprising a nucleic acid selected from the group
CC consisting of AEB86663, AEB86665, AEB86667, AEB86669, AEB86671, AEB86673,
CC AEB86675, AEB86677 and AEB86679; (2) an isolated, purified or enriched
CC nucleic acid for the production of a polyene polyketide; (3) an
CC expression vector comprising the novel nucleic acid or the nucleic acid
CC of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC 250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC the production of a polyketide. The nucleic acid can be used for
CC production of a polyketide, for use as a fungal and cancer cell growth
CC inhibitors. The present sequence represents a Streptomyces aizunensis
CC polyene polyketide synthase related contig 2 DNA sequence, which is used
CC in the exemplification of the present invention.
XX
SQ Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;

Query Match 12.9%; Score 129; DB 13; Length 164051;
Best Local Similarity 49.9%; Pred. No. 1.3e-11;
Matches 499; Conservative 0; Mismatches 480; Indels 22; Gaps 6;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||   ||  |  |||| ||   |    || ||||||   ||||||| |||
Db     111857 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 111916

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||   |||||||| | |||| ||           |  | |  | ||  |  || | |
Db     111917 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 111976

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              |||  |  | | ||  |  ||  | |  || |     | || ||     || |||  ||
Db     111977 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 112036

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225
                          || | || |||||||  |  |||| |   |  |   |||||||  |
Db     112037 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 112096

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285
              ||  |||    ||| |   || || || |   | || ||| | ||  | |||  |  ||
Db     112097 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 112156

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345
              | || ||   |||  | |   |  ||  ||  ||  |  ||  |    |       |
Db     112157 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 112216

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
              || |     || ||| |  |  | | |||||| | |      ||     |  |||||
Db     112217 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 112276

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              |  |  | | | |||||  ||||  || |  |  |||       | || ||| || | ||
Db     112277 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 112333

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
                | |  || ||          || |||  |    ||| ||  || ||   ||    |
Db     112334 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 112393

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
                | | | |  || |   | |  ||||| ||||    ||||  | ||||  |
Db     112394 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 112453

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
              |  |||  |   | ||||   |  | | |  | ||| | ||      | |    ||  |
Db     112454 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 112513

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
              ||| |    ||| ||| |   ||| || ||    |  || |  ||   | ||   || |
Db     112514 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 112573

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              ||| || || || || ||||||||| |||||||   |||  |||  | || |  |   |
Db     112574 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 112633

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
```

```
Db     112634 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 112691

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885

Db     112692 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 112751

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945

Db     112752 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 112810

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985

Db     112811 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 112851


RESULT 14
AAD54223
ID   AAD54223 standard; DNA; 24081 BP.
XX
AC   AAD54223;
XX
DT   17-JUN-2003  (first entry)
XX
DE   Streptomyces platensis subspecies rosaceus dorrigocin ORF6 DNA.
XX
KW   Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
KW   ds.
XX
OS   Streptomyces platensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .24081
FT                   /*tag=  a
FT                   /product= "ORF6 protein"
FT                   /note= "No start codon"
FT                   /partial
XX
PN   WO200288176-A2.
XX
PD   07-NOV-2002.
XX
PF   26-APR-2002; 2002WO-CA000591.
XX
PR   26-APR-2001; 2001US-0286346P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazopoulos E,  Staffa A,  Yang X;
XX
DR   WPI; 2003-201222/19.
DR   P-PSDB; AAE35489.
XX
PT   Novel isolated or purified polypeptide involved in biosynthesis of
PT   polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
PT   dorrigocin or lactimidomycin.
XX
PS   Claim 6; Page 168-181; 312pp; English.
XX
CC   The invention relates to novel proteins involved in the biosynthesis of
CC   polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC   microorganisms. Sequences of the invention allow direct manipulation of
CC   dorrigocin, lactimidomycin and related chemical structures via chemical
CC   engineering of the enzymes involved in the biosynthesis of dorrigocin and
CC   lactimidomycin. They are useful for introducing chemical handles into
CC   normally inert positions that permit subsequence chemical modifications
CC   and facilitate the development of polyketides. The genes and proteins of
CC   the invention can also be used to generate a focused library of analogues
CC   around a polyketide lead candidate to fine-tune the compound for optimal
CC   properties. They are useful for generating antibodies specific for the
CC   polyketide biosynthesis. The present sequence is Streptomyces platensis
CC   subspecies rosaceus DORR ORF6 DNA
XX
SQ   Sequence 24081 BP; 3269 A; 9645 C; 8186 G; 2981 T; 0 U; 0 Other;

  Query Match             12.6%;  Score 125.8;  DB 10;  Length 24081;
  Best Local Similarity   50.5%;  Pred. No. 5.5e-11;
  Matches  419;  Conservative    0;  Mismatches  392;  Indels   18;  Gaps    4;

Qy        173 CGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGT 232

Db     18354 CTCGCCGTCCTTCGCCGTCGAGCTGGCCGGCCGCGAGTTCTTCCTCGACGACCACCGGGT 18413

Qy        233 GCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGA 292

Db     18414 GCGCAACGTGCCGGTGCTCCCCGGCGTGGCCTATCTGGAGCTGGCGTACGCGGCGGCCCG 18473

Qy        293 GATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGT 352

Db     18474 GGCCGAGGGCGTCGACCCCGCCCACGCCCGCCTGCGCAACGTCGTCTGGTCGCGCCCCGC 18533

Qy        353 CGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGA 412

Db     18534 ACGGATCACCGGGCCGACCGCGGTCGAGATCGCGCTGCGGCCGTGCGAGGACGACGCCTT 18593

Qy        413 CTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCG 472

Db     18594 CACCTACGAGATCACCACGGCGGCCGACGGCGAACAGCCGGTGATCCACGGGCAAGGACG 18653

Qy        473 CCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCT 532

Db     18654 CATCGAGCGGTGCGGGACGCC------GTCACCCGCGCGCCTGGACATCGCCGCGCTGCG 18707

Qy        533 CTCACGTTGCTCGAGGGTCACCGGCGGA---GACGCGTGCTACCGCCGCCTCCAGCAGCT 589

Db     18708 CGCCCAGTGCGAGGTGCGCACTCTGGAACACGACGACTGCTACCGGCTCTTCGACCGCAT 18767

Qy        590 CGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGA 649

Db     18768 GGGCATCGGCTACGGCCCGGCCATGCGGGGCATCCGGCGGATCCACGTCGGCGCCGGGCT 18827

Qy        650 GGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCT 709

Db     18828 CGCCGTCGCACGCCTGAGCCTGCCGCAG---GCCGCCCGGGACGGCGCCGGCTGGGACCT 18884

Qy        710 CCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGC 769

Db     18885 GCACCCGTCGATGCTCGACGCCGCCGTACAGGCCACCTTGGGCCTGTCACTGGCCGAGGA 18944

Qy        770 GGCCCGAGCC------CCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCC 823

Db     18945 CACCGACACCGTGGCGCCGGCACTGCCCTTCGTCCTGGAGGAGGTGCAGCTGCTCGCGCC 19004

Qy        824 GCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGC 883

Db     19005 CAGCCCGGCCGGCGGGTGGGCCGTGGTGCGACCCGCAGCGGGCGACGGCGGCGGAGCCGT 19064

Qy        884 CAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGA 943

Db     19065 ACGCCGCATCGACATCGAACTGTGCGACGACGACGGCGAGGTGTGCGTACGCCTGCTCGG 19124

Qy        944 TTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992

Db     19125 GTTCACCGCACGCGTCCTGGCCGCCGGTGACGACCCCGCCGGCGGAGAG 19173


RESULT 15
AAD54217
ID   AAD54217 standard; DNA; 52101 BP.
XX
AC   AAD54217;
XX
DT   17-JUN-2003  (first entry)
XX
DE   Streptomyces platensis subspecies rosaceus dorrigocin DNA.
XX
```

```
KW   Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
KW   ds.
XX
OS   Streptomyces platensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(67. .3720)
FT                   /*tag= a
FT                   /product= "ORF1 protein"
FT   CDS             4092. .5681
FT                   /*tag= b
FT                   /product= "ORF2 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             5767. .6018
FT                   /*tag= c
FT                   /product= "ORF3 protein"
FT   CDS             6023. .7993
FT                   /*tag= d
FT                   /product= "ORF4 protein"
FT   CDS             8009. .17587
FT                   /*tag= e
FT                   /product= "ORF5 protein"
FT   CDS             17634. .41714
FT                   /*tag= f
FT                   /product= "ORF6 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             41772. .47633
FT                   /*tag= g
FT                   /product= "ORF7 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             47635. .49890
FT                   /*tag= h
FT                   /product= "ORF8 protein"
FT   CDS             49922. .50938
FT                   /*tag= i
FT                   /product= "ORF9 protein"
FT   CDS             51234. .52079
FT                   /*tag= j
FT                   /product= "ORF10 protein"
XX
PN   WO200288176-A2.
XX
PD   07-NOV-2002.
XX
PF   26-APR-2002; 2002WO-CA000591.
XX
PR   26-APR-2001; 2001US-0286346P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazopoulos E,  Staffa A,  Yang X;
XX
DR   WPI; 2003-201222/19.
DR   P-PSDB; AAE35484, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,
DR   AAE35490, AAE35491, AAE35492, AAE35493.
XX
PT   Novel isolated or purified polypeptide involved in biosynthesis of
PT   polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
PT   dorrigocin or lactimidomycin.
XX
PS   Claim 1; Page 85-113; 312pp; English.
XX
CC   The invention relates to novel proteins involved in the biosynthesis of
CC   polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC   microorganisms. Sequences of the invention allow direct manipulation of
CC   dorrigocin, lactimidomycin and related chemical structures via chemical
CC   engineering of the enzymes involved in the biosynthesis of dorrigocin and
CC   lactimidomycin. They are useful for introducing chemical handles into
CC   normally inert positions that permit subsequence chemical modifications
CC   and facilitate the development of polyketides. The genes and proteins of
CC   the invention can also be used to generate a focused library of analogues
CC   around a polyketide lead candidate to fine-tune the compound for optimal
CC   properties. They are useful for generating antibodies specific for the
CC   polyketide biosynthesis. The present sequence is Streptomyces platensis
CC   subspecies rosaceus DORR DNA
XX
SQ   Sequence 52101 BP; 7364 A; 20113 C; 17894 G; 6730 T; 0 U; 0 Other;

  Query Match             12.6%;  Score 125.8;  DB 10;  Length 52101;
  Best Local Similarity   50.5%;  Pred. No. 4.9e-11;
  Matches  419;  Conservative    0;  Mismatches  392;  Indels   18;  Gaps    4;

Qy        173 CGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGT 232
              | |||    | | | ||   |  |||     |||  || ||||| ||||| ||| ||||  ||
Db      35987 CTCGCCGTCCTTCGCCGTCGAGCTGGCCGGCCGCGAGTTCTTCCTCGACGACCACCGGGT 36046

Qy        233 GCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGA 292
              || |  |  |  | | |||||||  | ||| |  |||||| ||||   ||| || |
Db      36047 GCGCAACGTGCCGGTGCTCCCCGGCGTGGCCTATCTGGAGCTGGCGTACGCGGCGGCCCG 36106

Qy        293 GATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGT 352
              |  | |  ||    || | || || |  |||   | ||||        ||| |  ||
Db      36107 GGCCGAGGGCGTCGACCCCGCCCACGCCCGCCTGCGCAACGTCGTCTGGTCGCGCCCCGC 36166

Qy        353 CGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGA 412
                |  | ||| ||||  || | |||||||  |||| ||||||  ||  | || |  |
Db      36167 ACGGATCACCGGGCCGACCGCGGTCGAGATCGCGCTGCGGCCGTGCGAGGACGACGCCTT 36226

Qy        413 CTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCG 472
              |  |  || |  | ||| | ||| |     ||  |       |        || ||||     ||
Db      36227 CACCTACGAGATCACCACGGCGGCCGACGGCGAACAGCCGGTGATCCACGGGCAAGGACG 36286

Qy        473 CCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCT 532
              | ||||   | | |   |  |      | |  |  ||| |||      |||  |   | |
Db      36287 CATCGAGCGGTGCGGGACGCC------GTCACCCGCGCGCCTGGACATCGCCGCGCTGCG 36340

Qy        533 CTCACGTTGCTCGAGGGTCACCGGCGGA---GACGCGTGCTACCGCCGCCTCCAGCAGCT 589
              | | |  |||  |   |  |||     | |   ||||  |||||||| | | || | |   |
Db      36341 CGCCCAGTGCGAGGTGCGCACTCTGGAACACGACGACTGCTACCGGCTCTTCGACCGCAT 36400

Qy        590 CGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGA 649
              ||  |   | |||||     ||||||  |   |||     | | | |   ||||  ||
Db      36401 GGGCATCGGCTACGGCCCGGCCATGCGGGGCATCCGGCGGATCCACGTCGGCGCCGGGCT 36460

Qy        650 GGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCT 709
              |||  ||||     |  |||| ||| ||     |  |  ||||  | ||  ||  | |||
Db      36461 CGCCGTCGCACGCCTGAGCCTGCCGCAG---GCCGCCCGGGACGGCGCCGGCTGGGACCT 36517

Qy        710 CCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGC 769
              || ||  |  |||||||||| ||| | ||    |||  | || |||       |  ||||
Db      36518 GCACCCGTCGATGCTCGACGCCGCCGTACAGGCCACCTTGGGCCTGTCACTGGCCGAGGA 36577

Qy        770 GGCCCGAGCC------CCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCC 823
              ||     ||      ||    || || |||| | | |    |||   |||||||   ||
Db      36578 CACCGACACCGTGGCGCCGGCACTGCCCTTCGTCCTGGAGGAGGTGCAGCTGCTCGCGCC 36637

Qy        824 GCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGC 883
                |||||  |    | |||   |   ||  |  | |         |||     ||
Db      36638 CAGCCCGGCCGGCGGGTGGGCCGTGGTGCGACCCGCAGCGGGCGACGGCGGCGGAGCCGT 36697

Qy        884 CAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGA 943
              |    |    || ||   |||  |||    || ||||  ||||  | |      |||  |
Db      36698 ACGCCGCATCGACATCGAACTGTGCGACGACGACGGCGAGGTGTGCGTACGCCTGCTCGG 36757

Qy        944 TTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992
              |||||||  ||| ||  | | || |  | ||  ||| ||  ||   ||
Db      36758 GTTCACCGCACGCGTCCTGGCCGCCGGTGACGACCCCGCCGGCGGAGAG 36806


Search completed: April 27, 2006, 08:00:04
Job time : 488.212 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 206.315 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_35000_35999
Perfect score: 1000
Sequence: 1 tggactgggagccgctcgac.....acccccgcccagcggctcta 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	173.6	17.4	7680	3	US-09-902-540-8948	Sequence 8948, Ap
c	2	173.6	17.4	11476	3	US-09-902-540-955	Sequence 955, App
	3	162.4	16.2	5463	3	US-09-902-540-9150	Sequence 9150, Ap
c	4	162.4	16.2	9710	3	US-09-902-540-978	Sequence 978, App
	5	120	12.0	44377	2	US-08-804-227C-7	Sequence 7, Appli
	6	120	12.0	44377	2	US-08-804-198-1	Sequence 1, Appli
	7	114.6	11.5	23496	3	US-09-902-540-5645	Sequence 5645, Ap
c	8	114.6	11.5	51354	3	US-09-902-540-1270	Sequence 1270, Ap
	9	103	10.3	4893	3	US-09-902-540-8412	Sequence 8412, Ap
	10	103	10.3	15482	3	US-09-902-540-1067	Sequence 1067, Ap
	11	102.2	10.2	77536	3	US-09-410-551B-1	Sequence 1, Appli
	12	102.2	10.2	77536	3	US-09-940-316B-1	Sequence 1, Appli
	13	96.2	9.6	53799	3	US-10-042-665A-3	Sequence 3, Appli
	14	95.8	9.6	9005	3	US-09-902-540-7894	Sequence 7894, Ap
	15	95.8	9.6	10793	3	US-09-902-540-1062	Sequence 1062, Ap
	16	93.6	9.4	15872	3	US-09-105-537-1	Sequence 1, Appli
	17	93.6	9.4	15872	3	US-09-091-609-1	Sequence 1, Appli
	18	93.6	9.4	15872	3	US-09-091-609-3	Sequence 3, Appli
	19	90.6	9.1	5676	3	US-10-042-665A-1	Sequence 1, Appli
	20	86.6	8.7	6360	3	US-10-212-962-2	Sequence 2, Appli
c	21	81.6	8.2	1839	3	US-09-758-759-150	Sequence 150, App
c	22	81.6	8.2	109519	3	US-09-758-759-1	Sequence 1, Appli
	23	81.2	8.1	1584	3	US-09-902-540-5846	Sequence 5846, Ap
c	24	81.2	8.1	1585	3	US-09-902-540-135	Sequence 135, App
c	25	80.6	8.1	77536	3	US-09-410-551B-1	Sequence 1, Appli
c	26	80.6	8.1	77536	3	US-09-940-316B-1	Sequence 1, Appli
	27	80	8.0	28958	2	US-08-258-261B-6	Sequence 6, Appli
	28	80	8.0	28958	2	US-08-456-837-6	Sequence 6, Appli
	29	80	8.0	28958	2	US-08-457-342-6	Sequence 6, Appli
	30	80	8.0	28958	2	US-08-457-646A-6	Sequence 6, Appli
	31	80	8.0	28958	2	US-08-458-076A-6	Sequence 6, Appli
	32	80	8.0	28958	2	US-08-764-233A-4	Sequence 4, Appli
	33	80	8.0	28958	2	US-08-457-335A-6	Sequence 6, Appli
	34	80	8.0	28958	2	US-08-729-214-6	Sequence 6, Appli
	35	80	8.0	28958	3	US-09-028-934-6	Sequence 6, Appli
	36	80	8.0	49377	2	US-08-764-233A-1	Sequence 1, Appli
	37	79.8	8.0	2034	3	US-09-902-540-8949	Sequence 8949, Ap
	38	78	7.8	33529	3	US-09-144-085-3	Sequence 3, Appli
	39	77.6	7.8	68750	3	US-09-335-409-1	Sequence 1, Appli
	40	77.6	7.8	68750	3	US-09-568-102-1	Sequence 1, Appli
	41	77.6	7.8	68750	3	US-09-567-969-1	Sequence 1, Appli
	42	77.6	7.8	68750	3	US-09-568-480-1	Sequence 1, Appli
	43	77.6	7.8	68750	3	US-09-568-486-1	Sequence 1, Appli
	44	77.6	7.8	68750	3	US-09-568-472-1	Sequence 1, Appli
	45	77.6	7.8	68750	3	US-09-567-899-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-902-540-8948
; Sequence 8948, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8948
; LENGTH: 7680
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8948

Query Match 17.4%; Score 173.6; DB 3; Length 7680;
Best Local Similarity 52.5%; Pred. No. 1.2e-21;
Matches 532; Conservative 0; Mismatches 459; Indels 23; Gaps 6;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||| |  |||  | |   ||  |  ||  |||||       || ||||| || |
Db       2930 TCGACTGGGCGAAGCTGCATGGCGGCGCGCAGCTCCGGCGACGGAGCCTTCCCACCTATC 2989

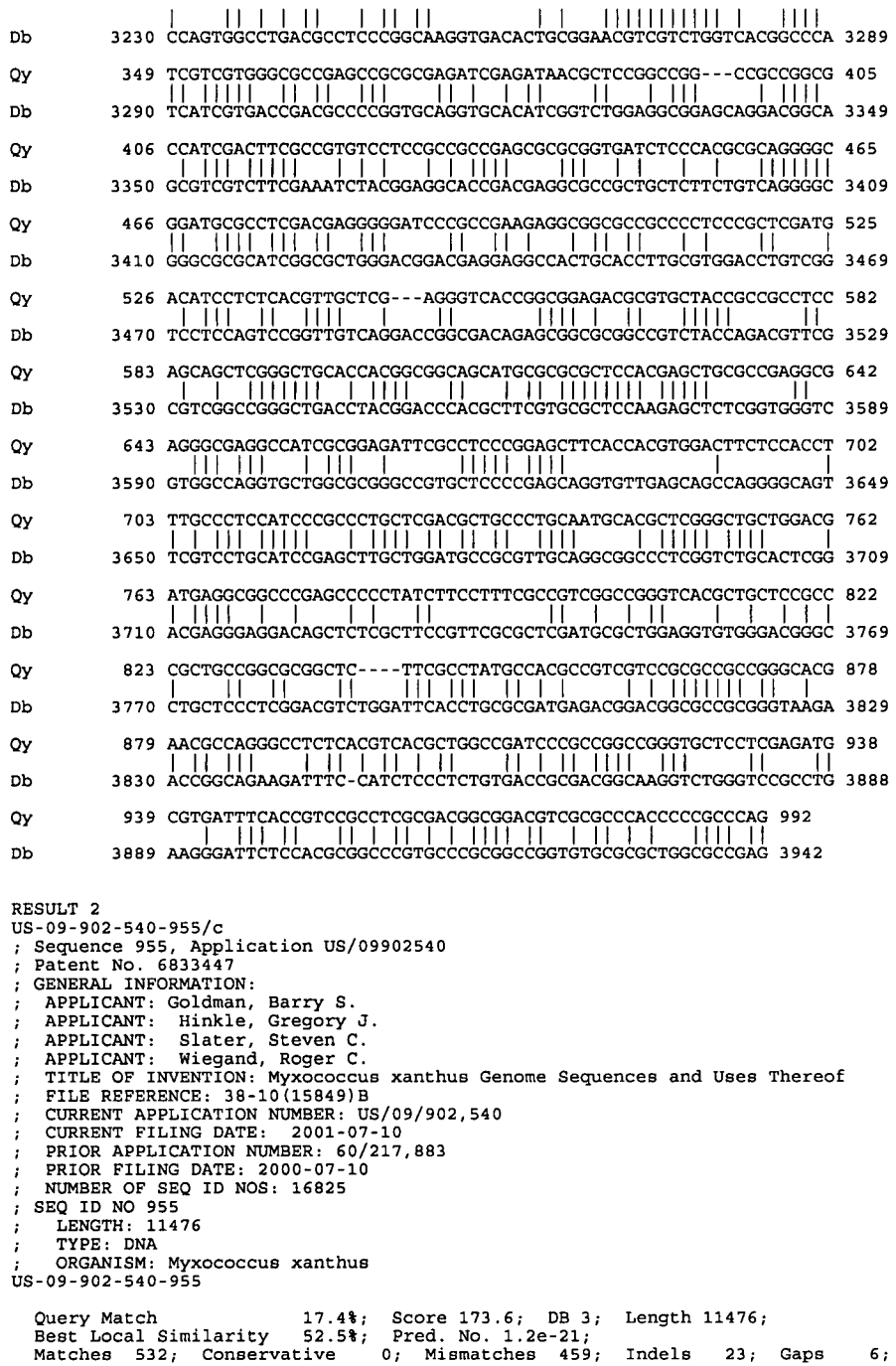
Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||||    |||   ||||||  |  | | |   |    |||    | || || || ||
Db       2990 CCTTCGCGAAGGAGAAGTACTGGGTAGCGAAGCCGGCGCGCGCGGAGAACCGCGTCCGGG 3049

Qy        121 CGGGAGAGAGGCG---CTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGC 177
              | ||      ||      | |||| ||||| ||||| | |||| | |||    | |  | ||
Db       3050 CCGGCAGCCGGGCGGCCCTGCATCCGCTGGTGCATGAGAACTCGTCCGACCTGCGGGGGC 3109

Qy        178 ACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACG 237
              |    |||  |  ||  |||  ||| || | ||||| ||||   || ||  |||||  |
Db       3110 AGCGCTACAGCGCCCACCTCACGGGAGAGGAGTTCTTCCTCCGAGACCATGTCGTGGGCC 3169

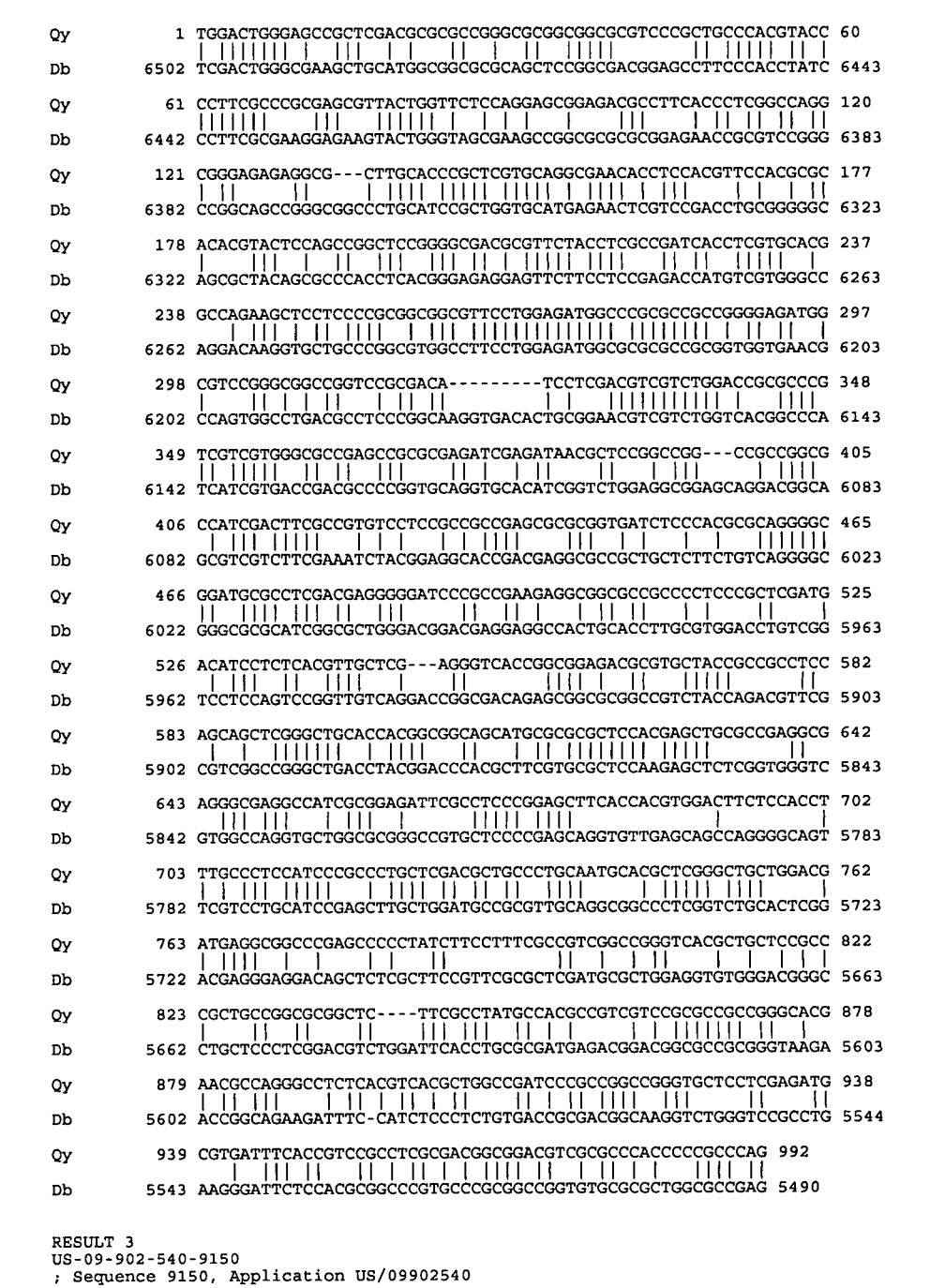
Qy        238 GCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGG 297
                | ||| | || ||||  | ||| ||||||||||||||| |||||||| | || ||  |
Db       3170 AGGACAAGGTGCTGCCCGGCGTGGCCTTCCTGGAGATGGCGCGCGCCGCGGTGGTGAACG 3229

Qy        298 CGTCCGGGCGGCCGGTCCGCGACA---------TCCTCGACGTCGTCTGGACCGCGCCCG 348
```



RESULT 2
US-09-902-540-955/c
; Sequence 955, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 955
; LENGTH: 11476
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-955

Query Match 17.4%; Score 173.6; DB 3; Length 11476;
Best Local Similarity 52.5%; Pred. No. 1.2e-21;
Matches 532; Conservative 0; Mismatches 459; Indels 23; Gaps 6;



RESULT 3
US-09-902-540-9150
; Sequence 9150, Application US/09902540

```
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9150
;   LENGTH: 5463
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-9150

  Query Match             16.2%;  Score 162.4;  DB 3;  Length 5463;
  Best Local Similarity   51.7%;  Pred. No. 1.1e-19;
  Matches  432;  Conservative    0;  Mismatches  386;  Indels   18;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||  || |||  |||||   |   |||  || ||  | |  || ||   |||||
Db       2909 TCGACTGGCGGCGGCTGTACGCGACGCACACGCCTCGTCGGCTGCGCCTCCCGGGGTACC 2968

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              | ||| |     |||||||||        ||  |   | |         | | |      |
Db       2969 CGTTCTCGAAGAAGCGTTACTCCCCGCCCGTGGTGGAGGCGCAGGCCGTCTCCACGGTGA 3028

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
               ||  |    |   | ||||| ||||| ||    |||| ||||    |   | |||
Db       3029 GAGGCGGCGCGGTGCTTCACCCCCTCGTCCACCGCAACAGCTCCGTCCTGGCTGAGCAGC 3088

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              |||  ||||  |||||||||||||| |||||||| || || || |  ||  |||| |
Db       3089 GCTACAGCAGCACGCTCCGGGGCGACGCCTTCTACCTGGCGGACCATCAGGTCAACGGGC 3148

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGC-----------GCCGCCG 288
               ||  |||| ||||| ||||| |  ||||||||||| |||           | |||
Db       3149 GCAAAGTCCTGCCCGCCGCGGCCTATCTGGAGATGGCTCGCTTCGCGGCCAGTGACGCAT 3208

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
              |||  |   |  | | | | | | | | | ||    || | |||||||| |||||
Db       3209 GGGCGGACGAAGCGGTGGGCCGGATACCTGTCCTCAGGAACATGGTCTGGACCTCGCCCT 3268

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
              | || ||||  | || | ||   |   |     |   || ||  || ||
Db       3269 TGGTGGTGGATGACGCGGCGGAGGTCGGAGTCGCGCTTCATGCGCAGGATGCGGGGCGGC 3328

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468
              |    ||||   |    |||||  ||||    ||||  |       |  ||  ||  | ||
Db       3329 TGACGTTCGAGATTCATTCCGCGCCCGACGCAGCGGGGGAGCGCACCGTCCATTGCCGCG 3388

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528
              |     ||   | ||| | ||  |      |        | ||  | ||| || |
Db       3389 GTCTGGCGGAGTTGCAGGAGCGCGTGGCGCCGCGCCATGACCTCCGGCGGCTGGAGGCGG 3448

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588
              | | |   ||      |||    || | ||   ||| |||     | ||   | ||
Db       3449 ATTGCACCCTGCGC------GTCCTGAGCCGGGAGCAGTGTTACGAGACCTTCGCCCGGC 3502

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648
              | ||  || || | ||   | |    || |||||    |  | ||||    ||  | |||
Db       3503 TTGGTTTGAACTATGGGCCCGGTCACCGTGCGCTTGTCCGGGTGCGGGTGGGGCACGGCC 3562

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708
              | ||  | |  ||| |    || || ||||  |     | |||  |    || | |  |
Db       3563 AAGCGCTGGGCGAGGTCGAGCTGCCCGAGCACCTGGTGGGCGACGGCCGATCGTTCGTGC 3622

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGG 768
              | || ||||| ||| ||||| | || ||||||      | ||||  || ||| ||| || |

Db       3623 TGCACCCCGCGCTGGTCGACTCCGCGCTGCAAGCGGTGGTCGGTGTGTTGGGCGACGAAG 3682

Qy         769 CGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCG 824
               | |   || |||   || || ||||| || |    ||||  | | |   |||||
Db       3683 GCGGCACCGCGCCCGCGCTGCCGTTCGCAGTGGAAGAGGTCGTGGTCCATGGCCCG 3738


RESULT 4
US-09-902-540-978/c
; Sequence 978, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 978
;   LENGTH: 9710
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-978

  Query Match             16.2%;  Score 162.4;  DB 3;  Length 9710;
  Best Local Similarity   51.7%;  Pred. No. 1.1e-19;
  Matches  432;  Conservative    0;  Mismatches  386;  Indels   18;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||  || |||  |||||   |   |||  || ||  | |  || ||   |||||
Db       2556 TCGACTGGCGGCGGCTGTACGCGACGCACACGCCTCGTCGGCTGCGCCTCCCGGGGTACC 2497

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              | ||| |     |||||||||        ||  |   | |         | | |      |
Db       2496 CGTTCTCGAAGAAGCGTTACTCCCCGCCCGTGGTGGAGGCGCAGGCCGTCTCCACGGTGA 2437

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
               ||  |    |   | ||||| ||||| ||    |||| ||||    |   | |||
Db       2436 GAGGCGGCGCGGTGCTTCACCCCCTCGTCCACCGCAACAGCTCCGTCCTGGCTGAGCAGC 2377

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              |||  ||||  |||||||||||||| |||||||| || || || |  ||  |||| |
Db       2376 GCTACAGCAGCACGCTCCGGGGCGACGCCTTCTACCTGGCGGACCATCAGGTCAACGGGC 2317

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGC-----------GCCGCCG 288
               ||  |||| ||||| ||||| |  ||||||||||| |||           | |||
Db       2316 GCAAAGTCCTGCCCGCCGCGGCCTATCTGGAGATGGCTCGCTTCGCGGCCAGTGACGCAT 2257

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
              |||  |   | | | | | | | | | ||    || | |||||||| |||||
Db       2256 GGGCGGACGAAGCGGTGGGCCGGATACCTGTCCTCAGGAACATGGTCTGGACCTCGCCCT 2197

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
              | || ||||  | || | ||   |   |     |   || ||  || ||
Db       2196 TGGTGGTGGATGACGCGGCGGAGGTCGGAGTCGCGCTTCATGCGCAGGATGCGGGGCGGC 2137

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468
              |    ||||   |    |||||  ||||    ||||  |       |  ||  ||  | ||
Db       2136 TGACGTTCGAGATTCATTCCGCGCCCGACGCAGCGGGGGAGCGCACCGTCCATTGCCGCG 2077

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528
              |     ||   | ||| | ||  |      |        | ||  | ||| || |
Db       2076 GTCTGGCGGAGTTGCAGGAGCGCGTGGCGCCGCGCCATGACCTCCGGCGGCTGGAGGCGG 2017

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588
              | | |   ||      |||    || | ||   ||| |||     | ||   | ||
Db       2016 ATTGCACCCTGCGC------GTCCTGAGCCGGGAGCAGTGTTACGAGACCTTCGCCCGGC 1963
```

```
Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648
Db       1962 TTGGTTTGAACTATGGGCCCGGTCACCGTGCGCTTGTCCGGGTGCGGGTGGGGCACGGCC 1903

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708
Db       1902 AAGCGCTGGGCGAGGTCGAGCTGCCCGAGCACCTGGTGGGCGACGGCCGATCGTTCGTGC 1843

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGG 768
Db       1842 TGCACCCCGCGCTGGTCGACTCCGCGCTGCAAGCGGTGGTCGGTGTGTTGGGCGACGAAG 1783

Qy        769 CGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCG 824
Db       1782 GCGGCACCGCGCCCGCGCTGCCGTTCGCAGTGGAAGAGGTCGTGGTCCATGGCCCG 1727
```

RESULT 5
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

Query Match 12.0%; Score 120; DB 2; Length 44377;
Best Local Similarity 48.1%; Pred. No. 2.1e-12;
Matches 450; Conservative 0; Mismatches 465; Indels 21; Gaps 3;

```
Qy         36 CGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCGCGAGCGTTACTGGT---------TC 86
Db      27376 CGCCAGGTGGAACTGCCCACGTACCCCTTCCAGCGCGAGCGCTACTGGTACAGCCCCGGC 27435

Qy         87 TCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTC 146
Db      27436 CACCGCGGTGACGACCCCGCCTCCCTCGGTCTGGACGCCGTCGACCACCCGCTGCTGGGC 27495

Qy        147 GTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGCCGGCTCCGGGGCGAC 206
Db      27496 AGCGGCGTCGAACTGCCGGAGTCCGGTGACCGGATGTACACCGCACGGCTGGGCGCCGAC 27555

Qy        207 GCGTTC---TACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCG 263
Db      27556 ACCACCCCGTGGCTGGCCGACCACGCGCTGCTGGGGTCGCCGCTGCTGCCCGGCGCCGCC 27615

Qy        264 TTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATC 323
Db      27616 TTCGCCGACCTGGCGCTCTGGGCCGGCCGCCAGGCCGGCACCGGCCGCGTCGAGGAGCTC 27675

Qy        324 CTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATA 383
Db      27676 ACCCTGGCCGCGCCCCTGGTGCTGCCCGGCTCCGGGGGTGTCCGGCTGCGGCTGAACGTC 27735

Qy        384 ACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCG 443
Db      27736 GGCGCCCCGGGCACCGACGACGCCCGCCGCTTCGCCGTGCACGCCCGCGCCGAGGGCGCC 27795

Qy        444 GTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCG 503
Db      27796 ACGGACTGGACCCTGCACGCCGAGGGGCTGCTCACCGCGCAGGACACGGCCGACGCGCCG 27855

Qy        504 GCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGAC 563
Db      27856 GACGCCTCG---------GCGGCCACCCCGCCCCCCGGCGCCGAACAACTGGACATCGGC 27906

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623
Db      27907 GACTTCTACCAGCGCTTCTCCGAACTCGGTTACGGCTACGGCCCGTTCTTCCGGGGACTG 27966

Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683
Db      27967 GTGAGCGCCCACCGCTGCGGCCCCGACATCCACGCGGAGGTCGCGCTGCCCGTCCAGGCG 28026

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743
Db      28027 CAGGGCGACGCGGCCCGCTTCGGCATCCATCCCGCGCTGCTGGACGCGGCGCTGCAGACC 28086

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803
Db      28087 ATGAGCCTCGGGGGCTTCTTCCCCGAGGACGGCCGCGTCCGCATGCCGTTCGCCCTGCGC 28146

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863
Db      28147 GGCGTTCGGCTGTACCGCGCCGGAGCCGACCGGCTGCACGTGCGCGTCTCGCCCGTCTCC 28206

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923
Db      28207 GAGGACGCGGTCCGCATCAGGTGCGCCGACGGCGAGGGACGGCCGGTCGCCGAGATCGAG 28266

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTC 959
Db      28267 TCCTTCATCATGCGGCCGGTCGACCCGGGACAGCTC 28302
```

RESULT 6
US-08-804-198-1
; Sequence 1, Application US/08804198

```
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             12.0%;  Score 120;  DB 2;  Length 44377;
  Best Local Similarity   48.1%;  Pred. No. 2.1e-12;
  Matches  450;  Conservative    0;  Mismatches  465;  Indels   21;  Gaps    3;

Qy         36 CGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCGCGAGCGTTACTGGT---------TC 86
              ||  |   ||    ||||||||||||||||||   |||||||| |||||||          |
Db      27376 CGCCAGGTGGAACTGCCCACGTACCCCTTCCAGCGCGAGCGCTACTGGTACAGCCCCGGC 27435

Qy         87 TCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTC 146
                | |  | |  ||| ||  | |||||||| | ||  |  |      | |  ||  | | |
Db      27436 CACCGCGGTGACGACCCCGCCTCCCTCGGTCTGGACGCCGTCGACCACCCGCTGCTGGGC 27495

Qy        147 GTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGCCGGCTCCGGGGCGAC 206
                 |   |    |   || |   |  | | |||| ||   |||||  | | ||||
Db      27496 AGCGGCGTCGAACTGCCGGAGTCCGGTGACCGGATGTACACCGCACGGCTGGGCGCCGAC 27555

Qy        207 GCGTTC---TACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCG 263
                |   |   |  || ||||| |||    |||  ||   |  ||| || ||||  || ||

Db      27556 ACCACCCCGTGGCTGGCCGACCACGCGCTGCTGGGGTCGCCGCTGCTGCCCGGCGCCGCC 27615

Qy        264 TTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATC 323
              |||   ||  ||||  | |   |||||       |||    |    | |  |||   ||  ||
Db      27616 TTCGCCGACCTGGCGCTCTGGGCCGGCCGCCAGGCCGGCACCGGCCGCGTCGAGGAGCTC 27675

Qy        324 CTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATA 383
                |   | ||         | ||    ||  |   ||  |  |  | ||    | |  |  |
Db      27676 ACCCTGGCCGCGCCCCTGGTGCTGCCCGGCTCCGGGGGTGTCCGGCTGCGGCTGAACGTC 27735

Qy        384 ACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCG 443
                 | |    ||| || ||||  |  ||||||||||  | ||  |||||| ||||
Db      27736 GGCGCCCCGGGCACCGACGACGCCCGCCGCTTCGCCGTGCACGCCCGCGCCGAGGGCGCC 27795

Qy        444 GTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCG 503
                |  ||   |  ||| |    |  ||  |||  || |  |||  | ||||| | | ||
Db      27796 ACGGACTGGACCCTGCACGCCGAGGGGCTGCTCACCGCGCAGGACACGGCCGACGCGCCG 27855

Qy        504 GCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGAC 563
                |   |  |              | | || ||   | |   || |           | |   | |
Db      27856 GACGCCTCG---------GCGGCCACCCCGCCCCCCGGCGCCGAACAACTGGACATCGGC 27906

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623
                | | |||||  ||| ||    | |||||       | |||||      | | || |  ||
Db      27907 GACTTCTACCAGCGCTTCTCCGAACTCGGTTACGGCTACGGCCCGTTCTTCCGGGGACTG 27966

Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683
                    | |||  |||     |||   |  |||||| |    || || |  |
Db      27967 GTGAGCGCCCACCGCTGCGGCCCCGACATCCACGCGGAGGTCGCGCTGCCCGTCCAGGCG 28026

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743
                || |  |||     ||  ||| | | |||||||||| ||||| ||||| || |||||   |
Db      28027 CAGGGCGACGCGGCCCGCTTCGGCATCCATCCCGCGCTGCTGGACGCGGCGCTGCAGACC 28086

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803
                | |  |    |    |    | || | |  |    |    | || ||||||  |  ||
Db      28087 ATGAGCCTCGGGGGCTTCTTCCCCGAGGACGGCCGCGTCCGCATGCCGTTCGCCCTGCGC 28146

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863
                | ||   ||||  |||| |     | |  |||||   ||      | | ||||    |||
Db      28147 GGCGTTCGGCTGTACCGCGCCGGAGCCGACCGGCTGCACGTGCGCGTCTCGCCCGTCTCC 28206

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923
                | |    ||| |   | |    | |||      ||      | | | |    |||    | |
Db      28207 GAGGACGCGGTCCGCATCAGGTGCGCCGACGGCGAGGGACGGCCGGTCGCCGAGATCGAG 28266

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTC 959
                || ||  |  ||  |    | | |  |  |||
Db      28267 TCCTTCATCATGCGGCCGGTCGACCCGGGACAGCTC 28302


RESULT 7
US-09-902-540-5645
; Sequence 5645, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5645
;    LENGTH: 23496
;    TYPE: DNA
;    ORGANISM: Myxococcus xanthus
US-09-902-540-5645
```

```
  Query Match             11.5%;  Score 114.6;  DB 3;  Length 23496;
  Best Local Similarity   48.0%;  Pred. No. 1.8e-11;
  Matches  410;  Conservative    0;  Mismatches  424;  Indels   21;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||  |   ||  ||| | |    |     |||| ||  ||    |||||  ||| |
Db      14882 TCGACTGGCGGAGTCTGCACGAGGGGAAGCGCAGGCGACGTCTCAGCTTGCCCCTGTATC 14941

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              | |||||     |||| |||  | |  |     |     | ||||  |  |   |      |
Db      14942 CGTTCGCGAAGCAGCGCTACGCGCTACCGGCCACGCCAACCGCCGTATCGGCC------G 14995

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
               ||    |  |||| ||||||||||  |||| | ||||   || ||| |  ||  |
Db      14996 TGGCTCCGGCGCGCCTGCACCCGCTGTTGCACGAGAACCAGTCGACGCTGGCCGGTCCGT 15055

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
               | |  ||  |   || | |||| |  || |||||| || ||| | || |  ||||
Db      15056 GCTTCGTCACGCACTTCAGCGGGGACGAGCCCTTCCTCGCGGACCACGTGGTCCGGGGCC 15115

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 300
              |||||  | ||  ||  | ||   |  | ||| ||  |||| | ||  ||  | |    | |
Db      15116 AGAAGGTGCTGCCTGGGGTCGGCTACCTCGAACTGGCGCACGAGGCGGTGCTGCGCTCAT 15175

Qy        301 CCGGGCGGCCGGTCCGCGAC--------------ATCCTCGACGTCGTCTGGACCGCGC 345
               |  | | |||  |||| |              |  | ||||| |||       |
Db      15176 TGGAACCGGCGGAACGCGGCTCGACGGCCATCCGTCTGGAGCAGGTCGTGTGGCTGCGTC 15235

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
              || |||||       || |||  |  |     |    ||  ||  | |||| ||
Db      15236 CCTTCGTCTACTCGGCGGAGAATGCGTCGCTGACCGTCCGACTGTCGCTTGTGGCCGACG 15295

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
               | |   | |  |  | | ||  | |||   |  | |      |     |||  |  |
Db      15296 AGAGCCTGTCCTTCCGGATCACCAGCACCGGCAGGTCCGGAGAGACGATGCTGCACGCCC 15355

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
               |  ||||     |   ||       | ||     |||| ||       |   |  |  |
Db      15356 AGGGGCGCGCGCGCCGTGGCCCGGGTGGCGTGATTGCGGAGCGCGTGGACATTCCAGGCG 15415

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
              |        |   | || |||       ||| || |||   ||| |  | |||||
Db      15416 CGCTGGCGGGCTGCACCCGGAAGGTGGTGAACGGCGAGGCGCTGTACGACGTCTTCCAGT 15475

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
               | ||  |  |  ||||   |       || | |  |  | ||  |||       || |||
Db      15476 CCATTGGCATCGATTACGGGCCCGCGCACCGGGTGATTGATGAACTGGCGGTGGGAGAGA 15535

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
               |  ||   ||||| || |    || |  | |     || ||  |||  | |  |||
Db      15536 ACCGGGTGGTCGCGAAGGTGTCGCTGCGCGGGACGAGCCTCGCAGACCGCGCTGCCTACA 15595

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              | || |  |||  |||||||||  |  |||||  ||   |||||| || |||  |   |
Db      15596 CGCTGCCGCCCAGCCTGCTCGATGGGGCCCTCCAGGCCACGCTGGGCCTGGCGATGGGAC 15655

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
              |||  || |    |  |||||  ||| || | | |    | | | |||||  ||
Db      15656 AGGGCGCGCTGAGAACGGCGCTTCCGTTCTCCCTGGACACCCTGCGGATTCTCCGTCCCT 15715

Qy        826 TGCCGGCGCGGCTCT 840
               ||||  | |  ||
Db      15716 GCCCGGACCAGGCCT 15730


RESULT 8
US-09-902-540-1270/c
; Sequence 1270, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1270
;   LENGTH: 51354
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(51354)
;   OTHER INFORMATION: unsure at all n locations
US-09-902-540-1270

  Query Match             11.5%;  Score 114.6;  DB 3;  Length 51354;
  Best Local Similarity   48.0%;  Pred. No. 1.8e-11;
  Matches  410;  Conservative    0;  Mismatches  424;  Indels   21;  Gaps    2;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||  |   ||  ||| | |    |     |||| ||  ||    |||||  ||| |
Db       9289 TCGACTGGCGGAGTCTGCACGAGGGGAAGCGCAGGCGACGTCTCAGCTTGCCCCTGTATC 9230

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              | |||||     |||| |||  | |  |     |     | ||||  |  |   |      |
Db       9229 CGTTCGCGAAGCAGCGCTACGCGCTACCGGCCACGCCAACCGCCGTATCGGCC------G 9176

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
               ||    |  |||| ||||||||||  |||| | ||||   || ||| |  ||  |
Db       9175 TGGCTCCGGCGCGCCTGCACCCGCTGTTGCACGAGAACCAGTCGACGCTGGCCGGTCCGT 9116

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
               | |  ||  |   || | |||| |  || |||||| || ||| | || |  ||||
Db       9115 GCTTCGTCACGCACTTCAGCGGGGACGAGCCCTTCCTCGCGGACCACGTGGTCCGGGGCC 9056

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 300
              |||||  | ||  ||  | ||   |  | ||| ||  |||| | ||  ||  | |    | |
Db       9055 AGAAGGTGCTGCCTGGGGTCGGCTACCTCGAACTGGCGCACGAGGCGGTGCTGCGCTCAT 8996

Qy        301 CCGGGCGGCCGGTCCGCGAC--------------ATCCTCGACGTCGTCTGGACCGCGC 345
               |  | | |||  |||| |              |  | ||||| |||       |
Db       8995 TGGAACCGGCGGAACGCGGCTCGACGGCCATCCGTCTGGAGCAGGTCGTGTGGCTGCGTC 8936

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
              || |||||       || |||  |  |     |    ||  ||  | |||| ||
Db       8935 CCTTCGTCTACTCGGCGGAGAATGCGTCGCTGACCGTCCGACTGTCGCTTGTGGCCGACG 8876

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
               | |   | |  |  | | ||  | |||   |  | |      |     |||  |  |
Db       8875 AGAGCCTGTCCTTCCGGATCACCAGCACCGGCAGGTCCGGAGAGACGATGCTGCACGCCC 8816

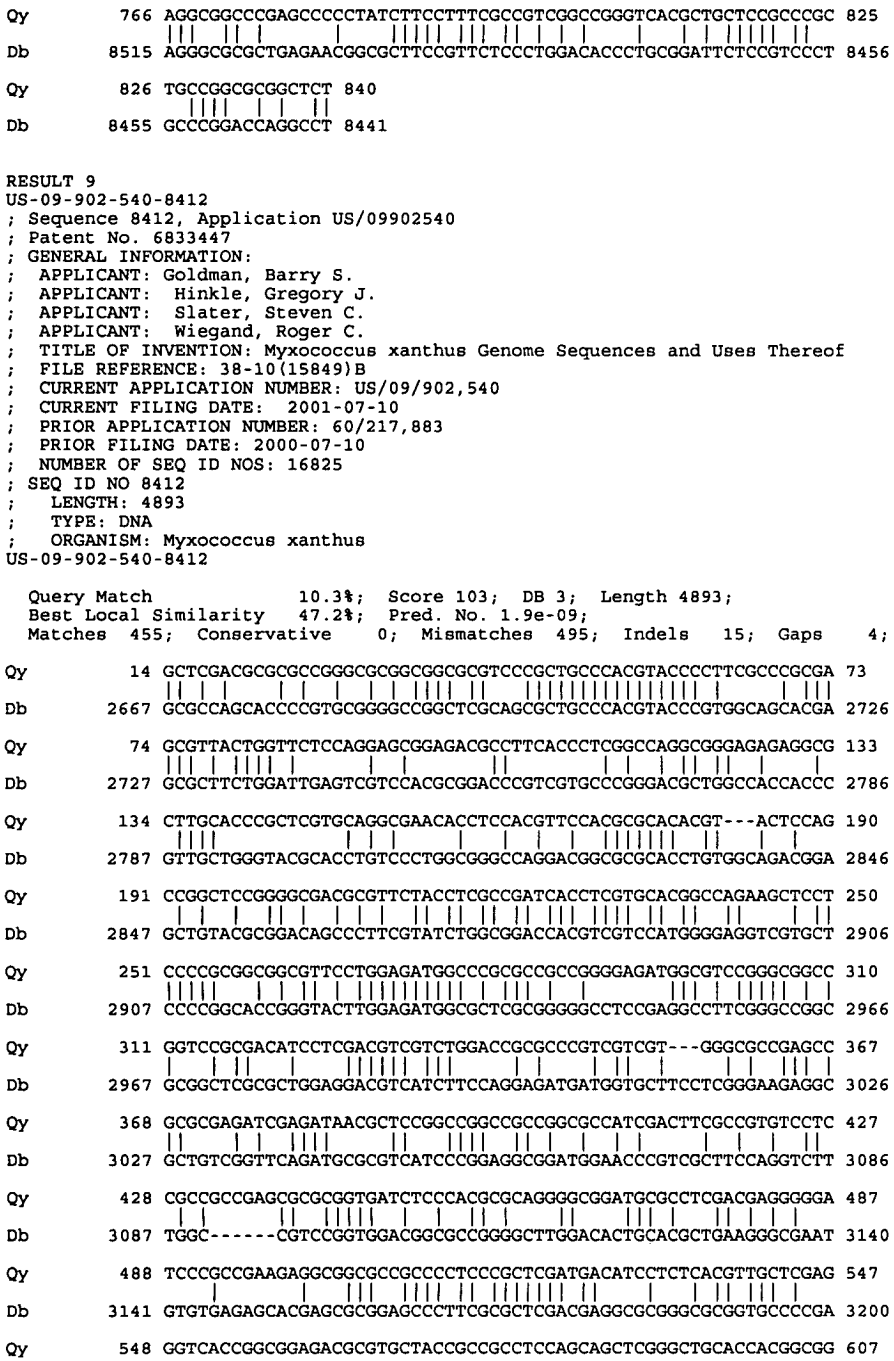
Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
               |  ||||     |   ||       | ||     |||| ||       |   |  |  |
Db       8815 AGGGGCGCGCGCGCCGTGGCCCGGGTGGCGTGATTGCGGAGCGCGTGGACATTCCAGGCG 8756

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
              |        |   | || |||       ||| || |||   ||| |  | |||||
Db       8755 CGCTGGCGGGCTGCACCCGGAAGGTGGTGAACGGCGAGGCGCTGTACGACGTCTTCCAGT 8696

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
               | ||  |  |  ||||   |       || | |  |  | ||  |||       || |||
Db       8695 CCATTGGCATCGATTACGGGCCCGCGCACCGGGTGATTGATGAACTGGCGGTGGGAGAGA 8636

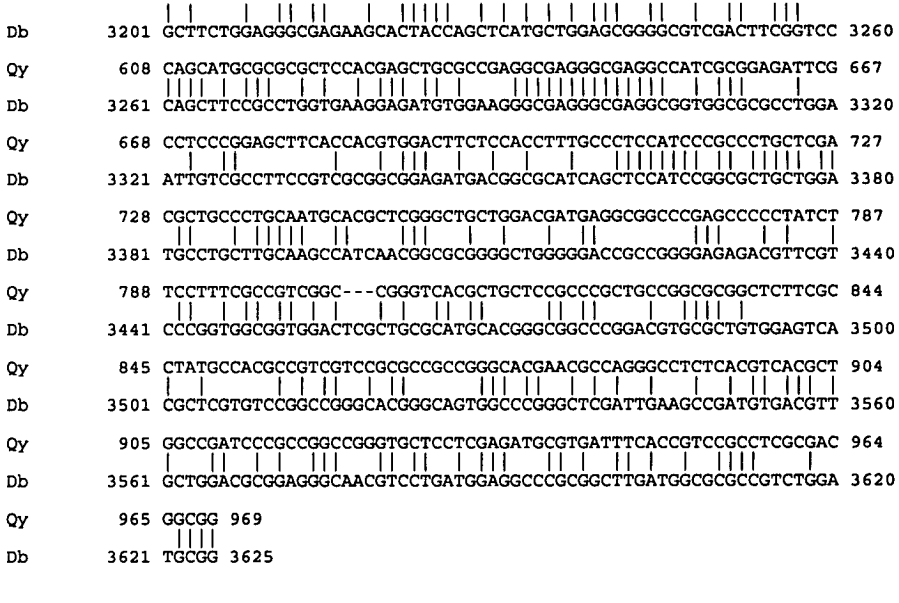
Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
               |  ||   ||||| || |    || |  | |     || ||  |||  | |  |||
Db       8635 ACCGGGTGGTCGCGAAGGTGTCGCTGCGCGGGACGAGCCTCGCAGACCGCGCTGCCTACA 8576

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
              | || |  |||  |||||||||  |  |||||  ||   |||||| || |||  |   |
Db       8575 CGCTGCCGCCCAGCCTGCTCGATGGGGCCCTCCAGGCCACGCTGGGCCTGGCGATGGGAC 8516
```



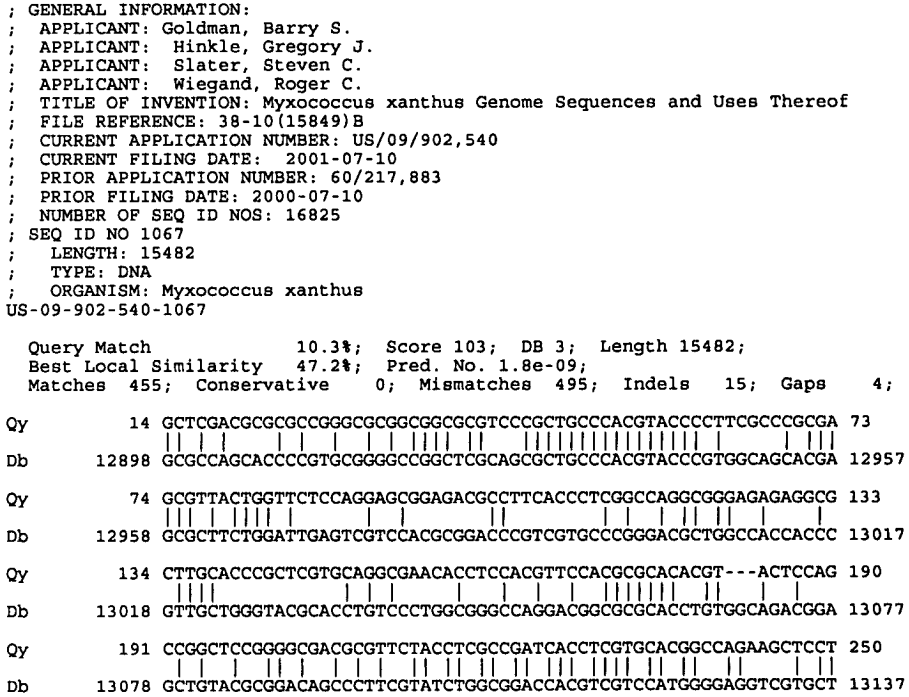
RESULT 9
US-09-902-540-8412
; Sequence 8412, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8412
; LENGTH: 4893
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8412

Query Match 10.3%; Score 103; DB 3; Length 4893;
Best Local Similarity 47.2%; Pred. No. 1.9e-09;
Matches 455; Conservative 0; Mismatches 495; Indels 15; Gaps 4;



RESULT 10
US-09-902-540-1067
; Sequence 1067, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1067
; LENGTH: 15482
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1067

Query Match 10.3%; Score 103; DB 3; Length 15482;
Best Local Similarity 47.2%; Pred. No. 1.8e-09;
Matches 455; Conservative 0; Mismatches 495; Indels 15; Gaps 4;



```
Qy        251 CCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCC 310
              ||||| | | || | |||||||||| | ||| | |       ||| | ||||| | |
Db      13138 CCCCGGCACCGGGTACTTGGAGATGGCGCTCGCGGGGGCCTCCGAGGCCTTCGGGCCGGC 13197

Qy        311 GGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGT---GGGCGCCGAGCC 367
              |   | ||   |   |||||| |||     | |     | ||  |     | |  ||| |
Db      13198 GCGGCTCGCGCTGGAGGACGTCATCTTCCAGGAGATGATGGTGCTTCCTCGGGAAGAGGC 13257

Qy        368 GCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTC 427
              ||    | |  ||||    ||   |||| || | | | | |     | |  |  ||
Db      13258 GCTGTCGGTTCAGATGCGCGTCATCCCGGAGGCGGATGGAACCCGTCGCTTCCAGGTCTT 13317

Qy        428 CGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGA 487
              | |      ||  ||||| |  | || |    ||    ||| |   || | | |
Db      13318 TGGC------CGTCCGGTGGACGGCGCCGGGGCTTGGACACTGCACGCTGAAGGGCGAAT 13371

Qy        488 TCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAG 547
                  |    |   |||  |||| || ||||||| ||     |   | || ||| |
Db      13372 GTGTGAGAGCACGAGCGCGGAGCCCTTCGCGCTCGACGAGGCGCGGGCGCGGTGCCCCGA 13431

Qy        548 GGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGG 607
              | |   |  || ||  | ||||  | | | | | |||  ||  |  ||  |||
Db      13432 GCTTCTGGAGGGCGAGAAGCACTACCAGCTCATGCTGGAGCGGGGCGTCGACTTCGGTCC 13491

Qy        608 CAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCG 667
              |||| | ||| | | | ||| || |    ||||||||||||||  | |||    |
Db      13492 CAGCTTCCGCCTGGTGAAGGAGATGTGGAAGGGCGAGGGCGAGGCGGTGGCGCGCCTGGA 13551

Qy        668 CCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGA 727
              |  ||       |   | |||  |  |  |   |  |||||||| || ||||| ||
Db      13552 ATTGTCGCCTTCCGTCGCGGCGGAGATGACGGCGCATCAGCTCCATCCGGCGCTGCTGGA 13611

Qy        728 CGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCT 787
              ||   | |||||  ||    |||   |  |   |  ||       |||   | |   |
Db      13612 TGCCTGCTTGCAAGCCATCAACGGCGCGGGGCTGGGGGACCGCCGGGGAGAGACGTTCGT 13671

Qy        788 TCCTTTCGCCGTCGGC---CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGC 844
              ||  | || || | |   | |  || ||      || ||     |  ||||
Db      13672 CCCGGTGGCGGTGGACTCGCTGCGCATGCACGGGCGGCCCGGACGTGCGCTGTGGAGTCA 13731

Qy        845 CTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCT 904
              | |      | | ||  | ||       ||| ||  | | |  |   |  | || ||| |
Db      13732 CGCTCGTGTCCGGCCGGGCACGGGCAGTGGCCCGGGCTCGATTGAAGCCGATGTGACGTT 13791

Qy        905 GGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGAC 964
              |   ||  | |  |||   || ||  | |||   || |  || |  |  ||||   |
Db      13792 GCTGGACGCGGAGGGCAACGTCCTGATGGAGGCCCGCGGCTTGATGGCGCGCCGTCTGGA 13851

Qy        965 GGCGG 969
              ||||
Db      13852 TGCGG 13856


RESULT 11
US-09-410-551B-1
; Sequence 1, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
;  APPLICANT: KOSAN BIOSCIENCES, Inc.
;  APPLICANT:  REEVES, CHRISTOPHER
;  APPLICANT:  CHU, DANIEL
;  APPLICANT:  KHOSLA, CHAITAN
;  APPLICANT:  SANTI, DANIEL
;  APPLICANT:  WU, KAI
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE: 30062-20026.00
;  CURRENT APPLICATION NUMBER: US/09/410,551B
;  CURRENT FILING DATE:  1999-10-01
;  PRIOR APPLICATION NUMBER: US 60/139,650
;  PRIOR FILING DATE: 1999-06-17
;  PRIOR APPLICATION NUMBER: US 60/123,810
;  PRIOR FILING DATE: 1999-03-11
;  PRIOR APPLICATION NUMBER: US 60/102,748
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 77536
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (52275)...(71465)
US-09-410-551B-1

  Query Match             10.2%;  Score 102.2;  DB 3;  Length 77536;
  Best Local Similarity   47.9%;  Pred. No. 2.4e-09;
  Matches  401;  Conservative    0;  Mismatches  418;  Indels   18;  Gaps    3;

Qy         11 GCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCG 70
              | |||||| |    || |||  ||   || |||| |||||||||||||| ||| |
Db      54641 GACGCTCGTCCTCGGCGGGGACCGCGCGCCCGTCACGCTGCCCACGTATCCGTTCCAGCA 54700

Qy         71 CGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAG 130
              | ||   |||||| |       | |      |||         ||||  || |   |||
Db      54701 CAAGGACTACTGGCTGCGGCCCACCTCCCGGGCCGATGTGACCGGCGCGGGGCAGGAGCA 54760

Qy        131 GCGCTTGCACCCGCT---------CGTGCAGGCGAACACCTCCACGTTCCACGCGCACAC 181
              |      |||||||||         || |   |||     ||  ||||  |   |    |
Db      54761 GGTGGCGCACCCGCTGCTCGGCGCCGCGGTCGCGCTGCCCGGCACGGGCGGAGTCGTCCT 54820

Qy        182 GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCA 241
              | |  | ||| |   | || | |     |  |||| ||| |||   ||  |||||
Db      54821 GACCGGCCGCCTGTCGCTGGCCTCCCATCCGTGGCTCGGCGAGCACGCGGTCGACGGCAC 54880

Qy        242 GAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGTC 301
              |||||| |||| |||| ||||| ||  | ||      ||||||  ||   ||
Db      54881 CGTGCTCCTGCCCGGCGCGGCCTTCCTCGAACTCGCGGCGCGCGCCGGCGACGAGGTCGG 54940

Qy        302 CGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGC 361
              | |    | | | ||| || ||  ||       | || ||    ||      ||||
Db      54941 CTGCGACCTGCTGCACGAACTCGTCATCGAGACGCCGCTCGTGCTGCCCGCGACCGGCGG 55000

Qy        362 CGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGT 421
              | | ||  |   |||||||  |      | | | ||    |    || ||||
Db      55001 TGTGGCGGTCTCCGTCGAGATCGCCGAACCCGACGACACGGGGCGGCGGGCGGTCACCGT 55060

Qy        422 GTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGA 481
              | |    |||||  || |||   | |     | || |   |     ||||| |
Db      55061 CCACGCGCGGGCCGACGGCTCGGGCCTGTGGACCCGACACGCCGGCGGATTCCTCGGCAC 55120

Qy        482 GGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTG 541
              ||      | | | |  ||| || | |   ||||             | | |
Db      55121 GG------CACCGGCACCGGCCACGGCCACGGACCCGGCACCCTGGCCGCCCGCGGAAGC 55174

Qy        542 CTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCA 601
              |     ||||  || ||  ||||    ||||  |||  || || |  |||||      | |
Db      55175 CGGACCGGTCGACGTCGCCGACG---TCTACGACCGGTTCGAGGACATCGGGTACTCCTA 55231

Qy        602 CGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGA 661
              |||      || | || | ||| |  |       | || | ||  | |   |   ||| ||
Db      55232 CGGACCGGGCTTCCGGGGGCTGCGGGCCGCCTGGCGCGCCGGCGACACCGTGTACGCCGA 55291

Qy        662 GATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCT 721
              | |    ||||| ||    ||  ||  |||  | ||   ||  | || || ||||| ||
Db      55292 GGTCGCGCTCCCCGACGAGCAGAGCGCCGACGCCGCCCGTTTCACGCTGCACCCCGCGCT 55351

Qy        722 GCTCGACGCTGCCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCC 781
              ||||||||| ||  | ||   |       |   | ||| |   | | | | || |
Db      55352 GCTCGACGCCGCGTTCCAGGCCGGCGCGCTGGCCGCGCTCGACGCACCCGGCGGGGCGGC 55411

Qy        782 CTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCT 838
              |   || || ||| |  ||       |||   | | |     |  || ||||||||
Db      55412 CCGACTGCCGTTCTCGTTCCAGGACGTCCGCATCCACGCGGCCGGGGCGACGCGGCT 55468
```

```
RESULT 12
US-09-940-316B-1
; Sequence 1, Application US/09940316B
; Patent No. 6759536
; GENERAL INFORMATION:
;  APPLICANT: KOSAN BIOSCIENCES, Inc.
;  APPLICANT:  REEVES, CHRISTOPHER
;  APPLICANT:  CHU, DANIEL
;  APPLICANT:  KHOSLA, CHAITAN
;  APPLICANT:  SANTI, DANIEL
;  APPLICANT:  WU, KAI
;  TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
;  TITLE OF INVENTION:  GENE CLUSTER
;  FILE REFERENCE: 30062-20026.11
;  CURRENT APPLICATION NUMBER: US/09/940,316B
;  CURRENT FILING DATE:  2001-08-27
;  PRIOR APPLICATION NUMBER: 09/410,551
;  PRIOR FILING DATE: 1999-10-01
;  PRIOR APPLICATION NUMBER: US 60/139,650
;  PRIOR FILING DATE: 1999-06-17
;  PRIOR APPLICATION NUMBER: US 60/123,810
;  PRIOR FILING DATE: 1999-03-11
;  PRIOR APPLICATION NUMBER: US 60/102,748
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 77536
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (52275)...(71465)
US-09-940-316B-1

  Query Match             10.2%;  Score 102.2;  DB 3;  Length 77536;
  Best Local Similarity   47.9%;  Pred. No. 2.4e-09;
  Matches  401;  Conservative    0;  Mismatches  418;  Indels   18;  Gaps    3;

Qy         11 GCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCG 70
              | ||||||  |    || |||  ||   || |||| ||||||||||||| || |||   |
Db      54641 GACGCTCGTCCTCGGCGGGGACCGCGCGCCCGTCACGCTGCCCACGTATCCGTTCCAGCA 54700

Qy         71 CGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAG 130
              | ||   ||||||  |       | |       |||         ||||  || |   |||
Db      54701 CAAGGACTACTGGCTGCGGCCCACCTCCCGGGCCGATGTGACCGGCGCGGGGCAGGAGCA 54760

Qy        131 GCGCTTGCACCCGCT---------CGTGCAGGCGAACACCTCCACGTTCCACGCGCACAC 181
              |     |||||||||         || |   |||    ||  ||||  |   |    |
Db      54761 GGTGGCGCACCCGCTGCTCGGCGCCGCGGTCGCGCTGCCCGGCACGGGCGGAGTCGTCCT 54820

Qy        182 GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCA 241
              |  |  | ||| |   | || | |       | |||| ||| |||   ||  |||||
Db      54821 GACCGGCCGCCTGTCGCTGGCCTCCCATCCGTGGCTCGGCGAGCACGCGGTCGACGGCAC 54880

Qy        242 GAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTC 301
              ||||| ||||  ||||| ||||| ||  | ||      |||||| ||   ||
Db      54881 CGTGCTCCTGCCCGGCGCGGCCTTCCTCGAACTCGCGGCGCGCCGGCGACGAGGTCGG 54940

Qy        302 CGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGC 361
              | |    | | | | |||  || ||  ||       |  || ||    ||    ||||
Db      54941 CTGCGACCTGCTGCACGAACTCGTCATCGAGACGCCGCTCGTGCTGCCCGCGACCGGCGG 55000

Qy        362 CGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGT 421
              | | ||  |    |||||||  |     |   | | | ||       |   || ||||
Db      55001 TGTGGCGGTCTCCGTCGAGATCGCCGAACCCGACGACACGGGGCGGCGGGCGGTCACCGT 55060

Qy        422 GTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGA 481
              | |    |||||  || |||   | ||    || | |   |  |||||  |
Db      55061 CCACGCGCGGGCCGACGGCTCGGGCCTGTGGACCCGACACGCCGGCGGATTCCTCGGCAC 55120

Qy        482 GGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTG 541
              ||       | | | |  |||  || | |   |||                 | | |
Db      55121 GG------CACCGGCACCGGCCACGGCCACGGACCCGGCACCCTGGCCGCCCGCGGAAGC 55174

Qy        542 CTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCA 601
              |     ||||  || ||  ||||    |||| |||  || ||| |  |||||      | |
Db      55175 CGGACCGGTCGACGTCGCCGACG---TCTACGACCGGTTCGAGGACATCGGGTACTCCTA 55231

Qy        602 CGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGA 661
              |||     || | || | ||| |  |      | || | ||  | |   |   ||| ||
Db      55232 CGGACCGGGCTTCCGGGGGCTGCGGGCCGCCTGGCGCGCCGGCGACACCGTGTACGCCGA 55291

Qy        662 GATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCT 721
              | |    ||||| ||     ||   ||  |||  | ||    ||  | || ||  |||| ||
Db      55292 GGTCGCGCTCCCCGACGAGCAGAGCGCCGACGCCGCCCGTTTCACGCTGCACCCCGCGCT 55351

Qy        722 GCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCC 781
              |||||||||  ||  | ||   |        |   ||| |   | | | | ||  |
Db      55352 GCTCGACGCCGCGTTCCAGGCCGGCGCGCTGGCCGCGCTCGACGCACCCGGCGGGGCGGC 55411

Qy        782 CTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCT 838
              |   || || ||| |  ||     |||    | | |     |  | || |||||||
Db      55412 CCGACTGCCGTTCTCGTTCCAGGACGTCCGCATCCACGCGGCCGGGGCGACGCGGCT 55468


RESULT 13
US-10-042-665A-3
; Sequence 3, Application US/10042665A
; Patent No. 6924106
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Toupet, Christine
;  APPLICANT:  Engel, Nathalie
;  TITLE OF INVENTION: Rifamycin biosynthesis gene cluster
;  FILE REFERENCE: 4-21001/B/C1
;  CURRENT APPLICATION NUMBER: US/10/042,665A
;  CURRENT FILING DATE:  2002-01-09
;  PRIOR APPLICATION NUMBER: 09/242,744
;  PRIOR FILING DATE: 1999-03-24
;  PRIOR APPLICATION NUMBER: PCT/EP97/04495
;  PRIOR FILING DATE: 1997-08-18
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 53799
;   TYPE: DNA
;   ORGANISM: Amycolatopsis mediterranei
US-10-042-665A-3

  Query Match             9.6%;  Score 96.2;  DB 3;  Length 53799;
  Best Local Similarity   48.6%;  Pred. No. 2.6e-08;
  Matches  365;  Conservative    0;  Mismatches  368;  Indels   18;  Gaps    3;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              ||  |  |||   |  ||  | ||  |   ||   || || |||  |||||| ||||||
Db      18203 TGACCGTGGACTGGTCCGGTGTGCTGCCACCGTCCCGCCGGGTCGAGCTGCCGACGTACG 18262

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGA---------GACGCCTTCACCC 111
              ||||||  || | |||  ||||||  |     | | ||||           ||||||  | | |
Db      18263 CCTTCGACCACCAGCACTACTGGCTGCAGATGGGCGGGTCGGCCACCGACGCCGTGTCGC 18322

Qy        112 TCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCC 171
              | |||| ||| || |   | |  ||  | |  || |   |   |  || |
Db      18323 TGGGCCTGGCCGGCGCCGACCACCCGCTGCTGGGCGCGGTCGTCCCGCTGCCGCAGTCCG 18382

Qy        172 ACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTC---TACCTCGCCGATCACC 228
              ||| || |   | | ||  |||||  |    |  ||  |   |  || ||||  |||
Db      18383 ACGGGCTCGTCTTCACCTCGCGGCTGTCGCTGAAGTCGCACCCGTGGCTGGCCGGGCACG 18442

Qy        229 TCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCG 288
              |   |||      ||||  | || |   ||| || | | ||  | || |    |||||
Db      18443 CGATCGGCGGGGTCGTGCTCATTCCGGGCACGGTGTACGTCGACCTCGCGCTGCGCGCCG 18502
```

```
Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
Db      18503 GCGACGAGCTCGGCTTCGGCGTCCTGGAAGAGCTCGTGATCGAGGCACCGCTGGTGCTGG 18562

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
Db      18563 GCGAGCGCGGCGGCGTTCGCGTGCAGGTCGCCGTGAGCGGGCCGAACGAGACCGGCTCGC 18622

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468
Db      18623 GTGCGGTGGACGTCTTCTCCATGCGGGAAGACGGCGACGAATGGACCCGGCACGCGACCG 18682

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528
Db      18683 GTCTCCTCGGGGCGTCGACGTCCCGGGAA------CCGAGCCGCTTCGACTTCGCCGCCT 18736

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588
Db      18737 GGCCGCCGGCCGGGGCGGAGCCGATCGACGTCGAAAACTTCTACACCGACCTCACCGAGC 18796

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648
Db      18797 GCGGGTACGCCTACAGCGGCGCCTTCCAGGGCATGCGGGCGGTCTGGCGGCGCGGTGACG 18856

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708
Db      18857 AGGTCTTCGCCGAGGTCGCGCTGCCTGACGACCACCGCGAGGACGCCGGCAAGTTCGGCC 18916

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCA 739
Db      18917 TCCACCCCGCCCTCCTCGACGCCGCTCTGCA 18947
```

RESULT 14
US-09-902-540-7894
; Sequence 7894, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7894
; LENGTH: 9005
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-7894

Query Match 9.6%; Score 95.8; DB 3; Length 9005;
Best Local Similarity 47.4%; Pred. No. 3.2e-08;
Matches 397; Conservative 0; Mismatches 422; Indels 18; Gaps 3;

```
Qy        182 GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCA 241
Db       5420 GGAATCCACACTGAGCGCGAGCGAGCCCCCGTTCCTCGCGGACCATCGCGTGGCGGGCTC 5479

Qy        242 GAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTC 301
Db       5480 GGTGGTGGTGCCGGGCGCGGCCTATCTGGAGATGGCGCTGTCCGCGGCGCATGCGATGTC 5539

Qy        302 CGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGC------GCCCGTCGTCGT 355
Db       5540 GGGAGGCACCGCGCACGAGCTGGTCGACACCTCCTTCAAGGAGGGGCTGCTCCTGCCCGA 5599

Qy        356 GGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTT 415
```

```
Db       5600 GGGCCAGGAGCGCACCGTCCAACTGGCGCTGCACGGCGAGCAGGGCCGCACCCTCACGTT 5659

Qy        416 CGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCT 475
Db       5660 TCAGCTGTCGAGCCAGAAGGCGGAAGCAGCGGGGTGGATGACGCACGTCGTGGGCAGGCT 5719

Qy        476 CGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTC 535
Db       5720 CCGCCTCTTGGACGCAGGGCAGGCGGGGCAGGCGTCCGAATCGTTGGAGGCCATCCGCGA 5779

Qy        536 ACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCT 595
Db       5780 GCGTTGCCCCGACGTCCTGACCCAGGCGGCGCATTACGAGACGCTGGCACGCAACGGCGT 5839

Qy        596 GCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCAT 655
Db       5840 GGCCTACGGGCCCGCGTTCCAGGGCGTGCAGCAGGTCTGGCGAGGGCGAGGGGAGGCGCT 5899

Qy        656 CGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCC 715
Db       5900 CGGCCACCTCCGCCTGCCGGAGTCATTGGCCGCTCGGGCCTCGGCGTATCGAATCCATCC 5959

Qy        716 CGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGC------ 769
Db       5960 CGCGTTGCTCGACGCCTGCTTCCAGTTGGTGTCCGCCGCGGTGCCTCCGGAGGACATGCC 6019

Qy        770 ---GGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCT 826
Db       6020 ACAGGACGTGGGGCCCGCCGTCCCGGTGGCGCTCGAGAACCTGCGGCTTCACGCACCGCC 6079

Qy        827 GCCGGCGCGGCTCTTCGCCTATGCC---ACGCCGTCGTCCGCGCCGCCGGGCACGAACGC 883
Db       6080 CCCGACGGAGGGCTTCTGCCACGTCCGGATGCGGACACTTGAGGGCACGCGCCGGGGCGT 6139

Qy        884 CAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGA 943
Db       6140 CTACGAGTTCGACCTCGTGCTGCGCGATGCGTCCGGCACGCTGCTGCTGGAGGCGCTGGG 6199

Qy        944 TTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 1000
Db       6200 GCTGCGGGTTCAGCAACTGGATGCTCCCGTGGCGGATGCGGAGGGCCGGGACCTCTA 6256
```

RESULT 15
US-09-902-540-1062
; Sequence 1062, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1062
; LENGTH: 10793
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1062

Query Match 9.6%; Score 95.8; DB 3; Length 10793;
Best Local Similarity 47.4%; Pred. No. 3.2e-08;
Matches 397; Conservative 0; Mismatches 422; Indels 18; Gaps 3;

```
Qy        182 GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCA 241
Db       5422 GGAATCCACACTGAGCGCGAGCGAGCCCCCGTTCCTCGCGGACCATCGCGTGGCGGGCTC 5481
```

```
Qy        242 GAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTC 301
              |  | |  | || |  ||||| |  |||||||||||| |   |||| | | |   | |||
Db       5482 GGTGGTGGTGCCGGGCGCGGCCTATCTGGAGATGGCGCTGTCCGCGGCGCATGCGATGTC 5541

Qy        302 CGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGC------GCCCGTCGTCGT 355
              ||  |  | |  | |||  |  |||||  |  ||  |  |        || | |   ||
Db       5542 GGGAGGCACCGCGCACGAGCTGGTCGACACCTCCTTCAAGGAGGGGCTGCTCCTGCCCGA 5601

Qy        356 GGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTT 415
              ||||   ||||    ||        |     || | |   |   | | || || ||  ||
Db       5602 GGGCCAGGAGCGCACCGTCCAACTGGCGCTGCACGGCGAGCAGGGCCGCACCCTCACGTT 5661

Qy        416 CGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCT 475
              |||| |      | |   || | |   |      |||| |    |    | ||
Db       5662 TCAGCTGTCGAGCCAGAAGGCGGAAGCAGCGGGGTGGATGACGCACGTCGTGGGCAGGCT 5721

Qy        476 CGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTC 535
              |  |     |||  | |   | | || |  | || ||    || | || | |||||  |
Db       5722 CCGCCTCTTGGACGCAGGGCAGGCGGGGCAGGCGTCCGAATCGTTGGAGGCCATCCGCGA 5781

Qy        536 ACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCT 595
              |||||| |    |||     |   |  |||   |||      ||    |    |||  |
Db       5782 GCGTTGCCCCGACGTCCTGACCCAGGCGGCGCATTACGAGACGCTGGCACGCAACGGCGT 5841

Qy        596 GCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCAT 655
              |  | ||||   |     | |  |   | ||  || |  | |||||    || |||||  |
Db       5842 GGCCTACGGGCCCGCGTTCCAGGGCGTGCAGCAGGTCTGGCGAGGGCGAGGGGAGGCGCT 5901

Qy        656 CGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCC 715
              ||   |  | ||||| ||||||         ||       |||  | | |    |||||||
Db       5902 CGGCCACCTCCGCCTGCCGGAGTCATTGGCCGCTCGGGCCTCGGCGTATCGAATCCATCC 5961

Qy        716 CGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGC------ 769
              |||  ||||||||||   | | || |    |  ||    | ||      ||||
Db       5962 CGCGTTGCTCGACGCCTGCTTCCAGTTGGTGTCCGCCGCGGTGCCTCCGGAGGACATGCC 6021

Qy        770 ---GGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCT 826
              || |   |  |||    | ||  | ||  |||       |  ||| | |   ||||
Db       6022 ACAGGACGTGGGGCCCGCCGTCCCGGTGGCGCTCGAGAACCTGCGGCTTCACGCACCGCC 6081

Qy        827 GCCGGCGCGGCTCTTCGCCTATGCC---ACGCCGTCGTCCGCGCCGCCGGGCACGAACGC 883
              ||| ||  |  ||||  | | | |   | || | |    | |    || ||  |  ||
Db       6082 CCCGACGGAGGGCTTCTGCCACGTCCGGATGCGGACACTTGAGGGCACGCGCCGGGGCGT 6141

Qy        884 CAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGA 943
              |   |  |   || ||  ||||  |||| |  ||||||  | |||| || |||  ||  |
Db       6142 CTACGAGTTCGACCTCGTGCTGCGCGATGCGTCCGGCACGCTGCTGCTGGAGGCGCTGGG 6201

Qy        944 TTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 1000
              |    || |  |    |   ||   ||| |||    |    | || |   |||||
Db       6202 GCTGCGGGTTCAGCAACTGGATGCTCCCGTGGCGGATGCGGAGGGCCGGGACCTCTA 6258
```

Search completed: April 27, 2006, 12:42:35
Job time : 208.315 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 1002.57 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_35000_35999
Perfect score: 1000
Sequence: 1 tggactgggagccgctcgac.....acccccgcccagcggctcta 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
	2	137.6	13.8	25085	6	US-10-132-134-35	Sequence 35, Appl
	3	137.6	13.8	50543	6	US-10-132-134-25	Sequence 25, Appl
	4	129	12.9	22533	9	US-10-760-493-36	Sequence 36, Appl
	5	129	12.9	164051	9	US-10-760-493-18	Sequence 18, Appl
	6	125.8	12.6	24081	6	US-10-132-134-13	Sequence 13, Appl
	7	125.8	12.6	52101	6	US-10-132-134-1	Sequence 1, Appli
	8	124.8	12.5	125401	7	US-10-203-295-35	Sequence 35, Appl
	9	115.2	11.5	65140	7	US-10-203-295-1	Sequence 1, Appli
	10	114.6	11.5	5766	9	US-10-937-379-48	Sequence 48, Appl
	11	114.6	11.5	73599	9	US-10-937-379-24	Sequence 24, Appl
c	12	109.2	10.9	138203	9	US-10-819-386A-1	Sequence 1, Appli
	13	105.2	10.5	15624	9	US-10-760-493-30	Sequence 30, Appl
	14	105	10.5	16299	9	US-10-760-493-32	Sequence 32, Appl
	15	103.8	10.4	64492	7	US-10-378-083-1	Sequence 1, Appli
	16	102.2	10.2	77536	3	US-09-940-316B-1	Sequence 1, Appli
	17	100.2	10.0	84428	7	US-10-229-148B-1	Sequence 1, Appli
	18	96.2	9.6	53799	9	US-10-042-665A-3	Sequence 3, Appli
	19	95.6	9.6	10287	9	US-10-760-493-24	Sequence 24, Appl
	20	95	9.5	29467	9	US-10-937-730A-1	Sequence 1, Appli
	21	95	9.5	67167	9	US-10-937-730A-3	Sequence 3, Appli
	22	94.4	9.4	14427	6	US-10-156-761-1540	Sequence 1540, Ap
c	23	94.4	9.4	9025608	6	US-10-156-761-1	Sequence 1, Appli
	24	93.8	9.4	9795	9	US-10-937-379-44	Sequence 44, Appl
	25	93.8	9.4	135638	6	US-10-314-657-1	Sequence 1, Appli
	26	93.8	9.4	135638	9	US-10-473-193-1	Sequence 1, Appli
	27	93.6	9.4	15872	3	US-09-861-289-1	Sequence 1, Appli
	28	93.6	9.4	15872	3	US-09-860-846-1	Sequence 1, Appli
	29	93.6	9.4	15872	3	US-09-988-384B-1	Sequence 1, Appli
	30	93.6	9.4	15872	3	US-09-836-821-1	Sequence 1, Appli
	31	93.6	9.4	15872	6	US-10-271-889-44	Sequence 44, Appl
c	32	93.4	9.3	88400	8	US-10-844-716-1	Sequence 1, Appli
	33	92.6	9.3	6297	6	US-10-132-134-37	Sequence 37, Appl
	34	90.6	9.1	5676	9	US-10-042-665A-1	Sequence 1, Appli
	35	90.2	9.0	9726	9	US-10-937-730A-2	Sequence 2, Appli
	36	88.6	8.9	23238	6	US-10-156-761-415	Sequence 415, App
	37	88.6	8.9	100000	6	US-10-156-761-15103	Sequence 15103, A
	38	87.8	8.8	11910	6	US-10-156-761-2879	Sequence 2879, Ap
c	39	87.8	8.8	125746	6	US-10-156-761-15102	Sequence 15102, A
	40	87.8	8.8	9025608	6	US-10-156-761-1	Sequence 1, Appli
	41	87.4	8.7	13416	5	US-10-205-032-11	Sequence 11, Appl
	42	87.4	8.7	60196	5	US-10-205-032-1	Sequence 1, Appli
	43	87.2	8.7	1023	7	US-10-437-963-75934	Sequence 75934, A
	44	86.8	8.7	17172	9	US-10-937-379-40	Sequence 40, Appl
	45	86.6	8.7	6360	5	US-10-212-962-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 1000; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 1.2e-231;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35000 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 35059

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35060 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 35119

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35120 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 35179
```

```
Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35180 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 35239

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35240 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGCGGGGAGATGGCGT 35299

Qy        301 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35300 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCG 35359

Qy        361 CCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35360 CCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCG 35419

Qy        421 TGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35420 TGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG 35479

Qy        481 AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35480 AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTT 35539

Qy        541 GCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35540 GCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACC 35599

Qy        601 ACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35600 ACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGG 35659

Qy        661 AGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35660 AGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCC 35719

Qy        721 TGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35720 TGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCC 35779

Qy        781 CCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35780 CCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCT 35839

Qy        841 TCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35840 TCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCA 35899

Qy        901 CGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35900 CGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCG 35959

Qy        961 CGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 1000
              ||||||||||||||||||||||||||||||||||||||||
Db      35960 CGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTCTA 35999
```

RESULT 2
US-10-132-134-35
; Sequence 35, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 35
; LENGTH: 25085
; TYPE: DNA
; ORGANISM: Streptomyces amphibiosporus
US-10-132-134-35

Query Match 13.8%; Score 137.6; DB 6; Length 25085;
Best Local Similarity 50.1%; Pred. No. 4.5e-24;
Matches 514; Conservative 0; Mismatches 464; Indels 48; Gaps 5;

```
Qy          3 GACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCC 62
              |||||| | ||||  |    |||  | ||  ||  || |  |||| ||||||||||
Db      23886 GACTGGCGGACGCTGTGGGGCGAACGGCACCCGCAACGGATCTCCATGCCGACGTACCCC 23945

Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
              ||| | | ||| |||||| ||  || || |       ||    | | ||||| |
Db      23946 TTCCAACTCCGGCGCTACTGGCTCGACATGACCA------CCCCGGCGCACGGCCCGCAC 23999

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |    |  | |||| ||||| |||||   ||||||||||    |   || |||
Db      24000 GTCTCCCGCGGACTGCATCCGCTGGTGCACCGGAACACCTCCGACCTGAGCGAGCAGCGC 24059

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              ||| ||   |  ||  |||  | ||||||| ||||||| |||| ||||| ||| |
Db      24060 TACACCTCGCACTTCACCGGCCGGGAGTTCTACATCGCCGACCACCGGGTGCAGGGCGAA 24119

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
              || || |||||  |||||||| |||||||||| ||||||||| |   | ||| |
Db      24120 CAGGTCGTCCCCGGCGCGGCGTTGCTGGAGATGGCACGCGCCGCCGCCGTCCTCGCGGCG 24179

Qy        303 GGGCGGCCGGTCCGCGAC------ATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTG 356
              ||  |  |||   ||||       | | | | || |||||| ||  |||  |   |
Db      24180 GGCGGTGCGGAGACCGACTGGGCGCTGCGCCAGGTGGTCTGGTCCCGGCCGCTGACGGCC 24239

Qy        357 GGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTC 416
              ||           || |  |         |  |  || ||| | |||||     | ||||
Db      24240 GGACGGCCCGTCGACGTGCACACCGCCGTGTCCGTGCGGGCGGACGGCGAACCGGCCTTC 24299

Qy        417 G---------CCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGG 467
              |         | | |  | ||| || ||| ||||  |||  ||||  ||  | |  ||
Db      24300 GAGATCTACACGGAGGGCCCCGGCGGCGAACGCGTCGTGCACTCCACCGGACGGCTGCAC 24359

Qy        468 ATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGAC 527
              |   || || |  |   |||        ||  |  ||    |||||   ||  | |
Db      24360 CGCAGGACCGCCGGGAACGCCGCCGAACTCCTGGACGGACCCGAACTCCCCGGCGGCGCC 24419

Qy        528 ATCCTCT-----------------------CACGTTGCTCGAGGGTCACCGGCGGAGAC 563
                | ||                        |   |  |  |||  ||  ||  ||
Db      24420 GGACACTTGGACGTCGCCGCGCTGCGTGCCCAGTGCGACGGCACCGTCCTCGACGCCGAG 24479

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623
              | |||||||  |||  ||        | || ||| |  ||||   ||   |||| ||  ||
Db      24480 GAGTGCTACGCCCGGTTCTCCGGCGTGGGCCTGGAGTACGGGCCCACGCTGCGAGCCGTC 24539

Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683
              |   |||| || | |||    |  |||| | ||    | | || ||  |||
Db      24540 GAGACGCTGAGCGGCGGCACCCGGCAGGCGGTGGCCCGGCTGCGGCTGTCGGCCGCCGCG 24599

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743
              || |        ||   || ||||| || ||   ||||||||||||| || || || | |
Db      24600 TCCGCGAGG---ACCGGGTTCGCCCTGCACCCGAGCCTGCTCGACGCCGCACTCCAGTCC 24656

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803
              |||   || ||  |   || |   |   |     |  |    || || || ||| | | |
Db      24657 ACGGCGGGCCTCTTCACCGGTTCCGGTACGTCCTCGGCGGCCCTGCCGTTTGCCCTGGAC 24716

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863
              ||| |   | |||| ||| ||  ||| |  |  |  ||   ||          |  ||
Db      24717 CGGCTGGAGGTGCTGCGCGCGACGCCCTCCTCGGGGTGGGCGGTGGCACGCTTCGCCGCC 24776

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923
              |     |         |   |   || ||    || |||   || ||||  |
Db      24777 GACGACCGCCCAGGCGGGGTGCGCCGCCTGGACATCGACGTGTGCGACGACGACGGCGAG 24836
```

```
Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACC 983
              |||  | |   ||| || |  |||   |||||          |  |||| ||| ||  ||
Db      24837 GTGTGCGTACGGATCCGGGGCTTCCAGGTCCGTACGTACGGCGGCGACGCCGCCCCGTCC 24896

Qy        984 CCCGCC 989
              || ||
Db      24897 GCCTCC 24902
```

RESULT 3
US-10-132-134-25
; Sequence 25, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 50543
; TYPE: DNA
; ORGANISM: Streptomyces amphibiosporus
US-10-132-134-25

Query Match 13.8%; Score 137.6; DB 6; Length 50543;
Best Local Similarity 50.1%; Pred. No. 3.9e-24;
Matches 514; Conservative 0; Mismatches 464; Indels 48; Gaps 5;

```
Qy          3 GACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCC 62
              ||||||  | ||||    |     |||   | ||  ||  || |  |||| |||||||||
Db      38432 GACTGGCGGACGCTGTGGGGCGAACGGCACCCGCAACGGATCTCCATGCCGACGTACCCC 38491

Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
              |||   | |  ||| |||||| ||  || || |       ||    | | ||||| |
Db      38492 TTCCAACTCCGGCGCTACTGGCTCGACATGACCA------CCCCGGCGCACGGCCCGCAC 38545

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |      |  |  |||| ||||| |||||   |||||||||     |   || |||
Db      38546 GTCTCCCGCGGACTGCATCCGCTGGTGCACCGGAACACCTCCGACCTGAGCGAGCAGCGC 38605

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              ||| ||   |   ||  |||   | |||||| ||||||| ||||  ||||| ||| |
Db      38606 TACACCTCGCACTTCACCGGCCGGGAGTTCTACATCGCCGACCACCGGGTGCAGGGCGAA 38665

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
              || || ||||||  |||||||| |||||||||| ||||||||||  |   | ||| |
Db      38666 CAGGTCGTCCCCGGCGCGGCGTTGCTGGAGATGGCACGCGCCGCCGCCGTCCTCGCGGCG 38725

Qy        303 GGGCGGCCGGTCCGCGAC------ATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTG 356
              ||  |  |||   ||||      | | | | || |||||| ||  |||  |    |
Db      38726 GGCGGTGCGGAGACCGACTGGGCGCTGCGCCAGGTGGTCTGGTCCCGGCCGCTGACGGCC 38785

Qy        357 GGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTC 416
              ||             || | |         |  |  || ||| | |||||     | ||||
Db      38786 GGACGGCCCGTCGACGTGCACACCGCCGTGTCCGTGCGGGCGGACGGCGAACCGGCCTTC 38845

Qy        417 G---------CCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGG 467
              |         | | |  | ||| || ||| ||||  |||  ||||  ||  | |  ||
Db      38846 GAGATCTACACGGAGGGCCCCGGCGGCGAACGCGTCGTGCACTCCACCGGACGGCTGCAC 38905

Qy        468 ATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGAC 527
              |   || || |   |   |||       ||  |  ||    |||||   ||  | |
Db      38906 CGCAGGACCGCCGGGAACGCCGCCGAACTCCTGGACGGACCCGAACTCCCCGGCGGCGCC 38965

Qy        528 ATCCTCT----------------------CACGTTGCTCGAGGGTCACCGGCGGAGAC 563
                | ||                      |   |   | |||  || ||  ||
Db      38966 GGACACTTGGACGTCGCCGCGCTGCGTGCCCAGTGCGACGGCACCGTCCTCGACGCCGAG 39025

Qy        564 GCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTC 623
              | |||||||  |||  ||        | || ||| |  ||||   ||   |||| ||  ||
Db      39026 GAGTGCTACGCCCGGTTCTCCGGCGTGGGCCTGGAGTACGGGCCCACGCTGCGAGCCGTC 39085

Qy        624 CACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCAC 683
              |   |||| || | |||    |  ||||  | ||   | | ||  ||  |||
Db      39086 GAGACGCTGAGCGGCGGCACCCGGCAGGCGGTGGCCCGGCTGCGGCTGTCGGCCGCCGCG 39145

Qy        684 CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGC 743
              || |        ||  || ||||| || ||  ||||||||||||| || || || | |
Db      39146 TCCGCGAGG---ACCGGGTTCGCCCTGCACCCGAGCCTGCTCGACGCCGCACTCCAGTCC 39202

Qy        744 ACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGC 803
              |||   || ||  |  || |  |  |  |  |  |  || || ||| | | |
Db      39203 ACGGCGGGCCTCTTCACCGGTTCCGGTACGTCCTCGGCGGCCCTGCCGTTTGCCCTGGAC 39262

Qy        804 CGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCC 863
              ||| |   | |||| ||| ||  |||  |  |  |  ||   ||        | ||
Db      39263 CGGCTGGAGGTGCTGCGCGCGACGCCCTCCTCGGGGTGGGCGGTGGCACGCTTCGCCGCC 39322

Qy        864 GCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGG 923
              |     |           |  |  |   || ||   ||  |||   || ||||  |
Db      39323 GACGACCGCCCAGGCGGGGTGCGCCGCCTGGACATCGACGTGTGCGACGACGACGGCGAG 39382

Qy        924 GTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACC 983
              |||  | |   ||| || |  |||   |||||          |  |||| ||| ||  ||
Db      39383 GTGTGCGTACGGATCCGGGGCTTCCAGGTCCGTACGTACGGCGGCGACGCCGCCCCGTCC 39442

Qy        984 CCCGCC 989
              || ||
Db      39443 GCCTCC 39448
```

RESULT 4
US-10-760-493-36
; Sequence 36, Application US/10760493
; Publication No. US20050187167A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia BioSciences Inc
; APPLICANT: Farnet, Chris M.
; APPLICANT: McAlpine, James B.
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Bachmann, Brian O.
; APPLICANT: Piraee, Mahmood
; TITLE OF INVENTION: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
; TITLE OF INVENTION: PHARMACEUTICALS
; FILE REFERENCE: 3004-9US
; CURRENT APPLICATION NUMBER: US/10/760,493
; CURRENT FILING DATE: 2004-01-21
; PRIOR APPLICATION NUMBER: USSN 60/441,123
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: USSN 60/469,810
; PRIOR FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: USSN 60/491,516
; PRIOR FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: USSN 60/494,568
; PRIOR FILING DATE: 2003-08-13
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 22533
; TYPE: DNA
; ORGANISM: Streptomyces aizunensis
US-10-760-493-36

Query Match 12.9%; Score 129; DB 9; Length 22533;
Best Local Similarity 49.9%; Pred. No. 5.3e-22;
Matches 499; Conservative 0; Mismatches 480; Indels 22; Gaps 6;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||    ||  |  |||| ||   |     || ||||||   |||||||| |||
```

```
Db       8069 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 8128

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120

Db       8129 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 8188

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180

Db       8189 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 8248

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225

Db       8249 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 8308

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285

Db       8309 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 8368

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345

Db       8369 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 8428

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405

Db       8429 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 8488

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465

Db       8489 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 8545

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525

Db       8546 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 8605

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585

Db       8606 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 8665

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645

Db       8666 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 8725

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705

Db       8726 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 8785

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765

Db       8786 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 8845

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825

Db       8846 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 8903

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885

Db       8904 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 8963

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945

Db       8964 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 9022

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985

Db       9023 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 9063


RESULT 5
US-10-760-493-18
; Sequence 18, Application US/10760493
; Publication No. US20050187167A1
; GENERAL INFORMATION:
;  APPLICANT: Ecopia BioSciences Inc
;  APPLICANT:  Farnet, Chris M.
;  APPLICANT:  McAlpine, James B.
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Bachmann, Brian O.
;  APPLICANT:  Piraee, Mahmood
;  TITLE OF INVENTION: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
;  TITLE OF INVENTION:  PHARMACEUTICALS
;  FILE REFERENCE: 3004-9US
;  CURRENT APPLICATION NUMBER: US/10/760,493
;  CURRENT FILING DATE:  2004-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/441,123
;  PRIOR FILING DATE: 2003-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/469,810
;  PRIOR FILING DATE: 2003-05-13
;  PRIOR APPLICATION NUMBER: USSN 60/491,516
;  PRIOR FILING DATE: 2003-08-01
;  PRIOR APPLICATION NUMBER: USSN 60/494,568
;  PRIOR FILING DATE: 2003-08-13
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 164051
;   TYPE: DNA
;   ORGANISM: Streptomyces aizunensis
US-10-760-493-18

  Query Match             12.9%;  Score 129;  DB 9;  Length 164051;
  Best Local Similarity   49.9%;  Pred. No. 3.7e-22;
  Matches  499;  Conservative    0;  Mismatches  480;  Indels   22;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60

Db     111857 TCGACTGGTCCGCGTACTTCGCGGGCACCGGCGCCCGCCGCGTCGAACTGCCCACCTACG 111916

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120

Db     111917 CCTTCCAGCGCGAGCGGTTCTGGCTCGAAGCCCCGGCCGGCTACATCGGCGACGTCGAAT 111976

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180

Db     111977 CGGCGGGCATGGGCGCGGCCCACCACCCGCTGCTCGGCGCCGCCGTCGCCCTCGCCGACG 112036

Qy        181 C------------GTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATC 225

Db     112037 GCGAAGGATTCCTGTTCACCGGCCGGCTCTCGCTCGACACCCACCCCTGGCTCGCCGACC 112096

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285

Db     112097 ACGCCGTCATGGGCAACGTCCTGCTGCCGGGCACCGCCTTCGTCGAACTCGCCATCCGCG 112156

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345

Db     112157 CGGGCGACCAGGCCGGCTGCGACCTCCTCGAAGAACTCACCCTCGAAGCACCGCTGATCC 112216

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405

Db     112217 TCGCCCCGCAGGCCGCGGCACGCCTCCAGATCGTGGTCGGAGCCCCCGACGGGTCCGGCC 112276

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465

Db     112277 GCCGCACCCTGGACGTGTACTCCAGCGACCCGGACGC---CCCCGCCGACGAGCCGTGGA 112333

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525

Db     112334 CCCGCCACGCCGGCGGCATCCTCGCCACCGGGGCACAGGCACCCGCCTTCGACCTGACCG 112393

Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585

Db     112394 CGTGGCCCCCGCCGGGCGCCGAAGCCGTCGGCGTCGACGGCCTCTACGAACACCTCGGCC 112453

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645

Db     112454 GGGGCGGCTTCGCCTACGGTCCCGTCTTCCAGGGGCTGCGCGCCGCCTGGCTCCTCGGCG 112513
```

```
Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
Db     112514 ACGACGTGTACGCCGAGGTCGCCCTGCCCGACGACCGGCAGGCCGAGGCCGCCCGGTTCG 112573

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
Db     112574 GCCTGCACCCGGCGCTCCTCGACGCGGCCCTGCACGCCACCTTCGTCCAGCCGTCCCCCG 112633

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
Db     112634 ACGGGGACCAGCAGGGCCGGCTGCCGTTCTCCTGGCGCGATGTGTCCCTGCAC--GCCGT 112691

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCA 885
Db     112692 CGGTGCGTCCGCGCTGCGCGTCCGCCTCACCCCCGACGGCCGGGACACCCTCTCCCTCCA 112751

Qy        886 GGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATT 945
Db     112752 -GCTCGCTGACACCACCGGCGCTCCCGTCGCCGCCGTCGGCCACCTGACGCTGCGGCCCG 112810

Qy        946 TCACCGTCCGCC-TCGCGACGGCGGACGTCGCGCCCACCCC 985
Db     112811 TCTCCGCCGACCAGCTCGGCAGCGCACGCTCCGCACACCAC 112851


RESULT 6
US-10-132-134-13
; Sequence 13, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Yang, Xianshu
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
;  FILE REFERENCE: 3012-2US
;  CURRENT APPLICATION NUMBER: US/10/132,134
;  CURRENT FILING DATE:  2002-04-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 24081
;   TYPE: DNA
;   ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-13

  Query Match             12.6%;  Score 125.8;  DB 6;  Length 24081;
  Best Local Similarity   50.5%;  Pred. No. 3.1e-21;
  Matches  419;  Conservative    0;  Mismatches  392;  Indels   18;  Gaps    4;

Qy        173 CGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGT 232
Db      18354 CTCGCCGTCCTTCGCCGTCGAGCTGGCCGGCCGCGAGTTCTTCCTCGACGACCACCGGGT 18413

Qy        233 GCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGA 292
Db      18414 GCGCAACGTGCCGGTGCTCCCCGGCGTGGCCTATCTGGAGCTGGCGTACGCGGCGGCCCG 18473

Qy        293 GATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGT 352
Db      18474 GGCCGAGGGCGTCGACCCCGCCCACGCCCGCCTGCGCAACGTCGTCTGGTCGCGCCCCGC 18533

Qy        353 CGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGA 412
Db      18534 ACGGATCACCGGGCCGACCGCGGTCGAGATCGCGCTGCGGCCGTGCGAGGACGACGCCTT 18593

Qy        413 CTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCG 472
Db      18594 CACCTACGAGATCACCACGGCGGCCGACGGCGAACAGCCGGTGATCCACGGGCAAGGACG 18653

Qy        473 CCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCT 532
Db      18654 CATCGAGCGGTGCGGGACGCC------GTCACCCGCGCGCCTGGACATCGCCGCGCTGCG 18707

Qy        533 CTCACGTTGCTCGAGGGTCACCGGCGGA---GACGCGTGCTACCGCCGCCTCCAGCAGCT 589
Db      18708 CGCCCAGTGCGAGGTGCGCACTCTGGAACACGACGACTGCTACCGGCTCTTCGACCGCAT 18767

Qy        590 CGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGA 649
Db      18768 GGGCATCGGCTACGGCCCGGCCATGCGGGGCATCCGGCGGATCCACGTCGGCGCCGGGCT 18827

Qy        650 GGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCT 709
Db      18828 CGCCGTCGCACGCCTGAGCCTGCCGCAG---GCCGCCCGGGACGGCGCCGGCTGGGACCT 18884

Qy        710 CCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGC 769
Db      18885 GCACCCGTCGATGCTCGACGCCGCCGTACAGGCCACCTTGGGCCTGTCACTGGCCGAGGA 18944

Qy        770 GGCCCGAGCC------CCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCC 823
Db      18945 CACCGACACCGTGGCGCCGGCACTGCCCTTCGTCCTGGAGGAGGTGCAGCTGCTCGCGCC 19004

Qy        824 GCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGC 883
Db      19005 CAGCCCGGCCGGCGGGTGGGCCGTGGTGCGACCCGCAGCGGGCGACGGCGGCGGAGCCGT 19064

Qy        884 CAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGA 943
Db      19065 ACGCCGCATCGACATCGAACTGTGCGACGACGACGGCGAGGTGTGCGTACGCCTGCTCGG 19124

Qy        944 TTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992
Db      19125 GTTCACCGCACGCGTCCTGGCCGCCGGTGACGACCCCGCCGGCGGAGAG 19173


RESULT 7
US-10-132-134-1
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Yang, Xianshu
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
;  FILE REFERENCE: 3012-2US
;  CURRENT APPLICATION NUMBER: US/10/132,134
;  CURRENT FILING DATE:  2002-04-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 52101
;   TYPE: DNA
;   ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-1

  Query Match             12.6%;  Score 125.8;  DB 6;  Length 52101;
  Best Local Similarity   50.5%;  Pred. No. 2.7e-21;
  Matches  419;  Conservative    0;  Mismatches  392;  Indels   18;  Gaps    4;

Qy        173 CGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGT 232
Db      35987 CTCGCCGTCCTTCGCCGTCGAGCTGGCCGGCCGCGAGTTCTTCCTCGACGACCACCGGGT 36046

Qy        233 GCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGA 292
Db      36047 GCGCAACGTGCCGGTGCTCCCCGGCGTGGCCTATCTGGAGCTGGCGTACGCGGCGGCCCG 36106

Qy        293 GATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGT 352
Db      36107 GGCCGAGGGCGTCGACCCCGCCCACGCCCGCCTGCGCAACGTCGTCTGGTCGCGCCCCGC 36166

Qy        353 CGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGA 412
```

```
Db      36167 ACGGATCACCGGGCCGACCGCGGTCGAGATCGCGCTGCGGCCGTGCGAGGACGACGCCTT 36226

Qy        413 CTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCG 472

Db      36227 CACCTACGAGATCACCACGGCGGCCGACGGCGAACAGCCGGTGATCCACGGGCAAGGACG 36286

Qy        473 CCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCT 532

Db      36287 CATCGAGCGGTGCGGGACGCC------GTCACCCGCGCGCCTGGACATCGCCGCGCTGCG 36340

Qy        533 CTCACGTTGCTCGAGGGTCACCGGCGGA---GACGCGTGCTACCGCCGCCTCCAGCAGCT 589

Db      36341 CGCCCAGTGCGAGGTGCGCACTCTGGAACACGACGACTGCTACCGGCTCTTCGACCGCAT 36400

Qy        590 CGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGA 649

Db      36401 GGGCATCGGCTACGGCCCGGCCATGCGGGGCATCCGGCGGATCCACGTCGGCGCCGGGCT 36460

Qy        650 GGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCT 709

Db      36461 CGCCGTCGCACGCCTGAGCCTGCCGCAG---GCCGCCCGGGACGGCGCCGGCTGGGACCT 36517

Qy        710 CCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGC 769

Db      36518 GCACCCGTCGATGCTCGACGCCGCCGTACAGGCCACCTTGGGCCTGTCACTGGCCGAGGA 36577

Qy        770 GGCCCGAGCC------CCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCC 823

Db      36578 CACCGACACCGTGGCGCCGGCACTGCCCTTCGTCCTGGAGGAGGTGCAGCTGCTCGCGCC 36637

Qy        824 GCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGC 883

Db      36638 CAGCCCGGCCGGCGGGTGGGCCGTGGTGCGACCCGCAGCGGGCGACGGCGGCGGAGCCGT 36697

Qy        884 CAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGA 943

Db      36698 ACGCCGCATCGACATCGAACTGTGCGACGACGACGGCGAGGTGTGCGTACGCCTGCTCGG 36757

Qy        944 TTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAG 992

Db      36758 GTTCACCGCACGCGTCCTGGCCGCCGGTGACGACCCCGCCGGCGGAGAG 36806


RESULT 8
US-10-203-295-35
; Sequence 35, Application US/10203295
; Publication No. US20040115762A1
; GENERAL INFORMATION:
;  APPLICANT: Zotchev, Sergey Borisovich
;  APPLICANT:  Sekurova, Olga Nikalayivna
;  APPLICANT:  Fjaervik, Epsen
;  APPLICANT:  Brautaset, Trygve
;  APPLICANT:  Strom, Arne Reidar
;  APPLICANT:  Valla, Svein
;  APPLICANT:  Ellingsen, Trond Erling
;  APPLICANT:  Sletta, Havard
;  APPLICANT:  Gulliksen, Ole-Martin
;  TITLE OF INVENTION: Novel genes encoding a nystatin polyketide synthase and their
;  TITLE OF INVENTION:  manipulation and utility
;  FILE REFERENCE: 1181-265
;  CURRENT APPLICATION NUMBER: US/10/203,295
;  CURRENT FILING DATE:  2003-05-19
;  PRIOR APPLICATION NUMBER: PCT/GB 01/00509
;  PRIOR FILING DATE: 2001-02-08
;  PRIOR APPLICATION NUMBER: GB 0002840.7
;  PRIOR FILING DATE: 2000-02-08
;  PRIOR APPLICATION NUMBER: GB 0008786.6
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: GB 0009387.2
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.0

; SEQ ID NO 35
;   LENGTH: 125401
;   TYPE: DNA
;   ORGANISM: Streptomyces noursei ATCC 11455
US-10-203-295-35

  Query Match           12.5%;  Score 124.8;  DB 7;  Length 125401;
  Best Local Similarity   50.8%;  Pred. No. 4e-21;
  Matches  467;  Conservative    0;  Mismatches  427;  Indels   26;  Gaps    6;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60

Db      37451 TGGACTGGACGGCGTTCTACGCCGGCACCGGCGCCCACCGCACCGACCTGCCGACCTACG 37510

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTT---------CTCCAGGAGCGGAGACGCCTTCACCC 111

Db      37511 CCTTCCAGTACGAGCGCTACTGGCCCAAGGCCACCTACCGGCCCGCCGACGCCACCGGCC 37570

Qy        112 TCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCC 171

Db      37571 TCGGCCTGACCGCCGCCGACCACCCGCTGCTCGGCGCCGCCATGTCCGTCGCCGGGTCCG 37630

Qy        172 ACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATCACC 228

Db      37631 ACGAGCTCCTGCTCACCGGCACCCTGTCGCTCGCCACCCACCCCTGGCTCGCCGACCACG 37690

Qy        229 TCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCG 288

Db      37691 TCGTCGGCGGCATGGTCTTCTTCCCCGGCACCGGCTTCCTGGAACTGGCGGTCCGCGCCG 37750

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348

Db      37751 CCGACCAGGTCGGCTGCGACCGGGTCGAGGAACTCATGCTCGCCG------CGCCGCTGA 37804

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408

Db      37805 TCCTGCCCGCCACCGGCACCGTCCAGATGCAGATCGCGGTCGGCGCCGCGGACGACGACG 37864

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468

Db      37865 GCGGCCGCGACCTGCGCTTCTTCACCCGGCCCGGGGACGACCCGGACGCCGCCTGGGCCC 37924

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528

Db      37925 AGCACGCCACCGGCCGGATCACCGAGGGCGAGCGCGTCCTCGCCCTCGACACCACCACCT 37984

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588

Db      37985 GGCCGCCCCGCGACGCGGAACCCGTCGACATCGACGGCCTCTACGACCGCTACCGCGCCA 38044

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648

Db      38045 ACGGACTCGACTACGGGCCCGTCTTCCGCGGCCTGCGCGCCGTATGGCGCCGCGACACCG 38104

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708

Db      38105 AGATCTACGCCGAGGTCGCCCTGCCCGAAG---GCACCGCCGACGCCGACGCCTTCGGCC 38161

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGG 768

Db      38162 TGCACCCGGCCCTCTTCGACGCCGTCCTGCACAGCACCCTCTTCGCCTCCGCCGACGGCG 38221

Qy        769 CGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGC 828

Db      38222 ACGACCGCAGCCTC---CTGCCGTTCGCCTGGAACGGCGTGTCCCTGCAC--GCCGCGGG 38276

Qy        829 CGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGG 888

Db      38277 CGCGGACGCGCTGCGCGTCCGGATCACCAGCTGCGGCCCCGACGCCGTGGAGATCACCGC 38336

Qy        889 CCTCTCACGTCACGCTGGCC 908

Db      38337 CGTCGACCCGCAGGGCCGCC 38356
```

RESULT 9
US-10-203-295-1
; Sequence 1, Application US/10203295
; Publication No. US20040115762A1
; GENERAL INFORMATION:
; APPLICANT: Zotchev, Sergey Borisovich
; APPLICANT: Sekurova, Olga Nikalayivna
; APPLICANT: Fjaervik, Epsen
; APPLICANT: Brautaset, Trygve
; APPLICANT: Strom, Arne Reidar
; APPLICANT: Valla, Svein
; APPLICANT: Ellingsen, Trond Erling
; APPLICANT: Sletta, Havard
; APPLICANT: Gulliksen, Ole-Martin
; TITLE OF INVENTION: Novel genes encoding a nystatin polyketide synthase and their
; TITLE OF INVENTION: manipulation and utility
; FILE REFERENCE: 1181-265
; CURRENT APPLICATION NUMBER: US/10/203,295
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/GB 01/00509
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: GB 0002840.7
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: GB 0008786.6
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: GB 0009387.2
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 65140
; TYPE: DNA
; ORGANISM: Streptomyces noursei ATCC 11455
US-10-203-295-1

Query Match 11.5%; Score 115.2; DB 7; Length 65140;
Best Local Similarity 48.1%; Pred. No. 9.2e-19;
Matches 429; Conservative 0; Mismatches 448; Indels 15; Gaps 3;

```
Qy        5 CTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTT 64
Db    19120 CTGGGCCGGCCTCTTCGACGGCACCGGCGCGCGCCGCGCCGACCTGCCCACCTACCCCTT 19179

Qy       65 CGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGG 124
Db    19180 CCAGCACCAGCGGTTCTGGCCGACCGCGGCCCGCGCCGCCCAGGACGTCACCGCCGCGGG 19239

Qy      125 AGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGC------- 177
Db    19240 ACTGGGCGCCGCCGACCACCCGCTGCTCGGCGCCACCGTCGAACTCGCCGACGGGGCCGG 19299

Qy      178 --ACACGTACTCCAGCCGGCTCCGGGGCG---ACGCGTTCTACCTCGCCGATCACCTCGT 232
Db    19300 CTACTTGTTCACCAGCCGGCTCTCCGTCCGGACCCACCCCTGGCTCGCCGACCACGGGGT 19359

Qy      233 GCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGA 292
Db    19360 CCAGGGCCGGGCCCTGCTGCCCGGCACCGCCTTCGTCGAACTGGCCGTCCGCGCCGGCGA 19419

Qy      293 GATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGT 352
Db    19420 CGAGGCCGGCTGCGACCGCGTCGAGGAACTGACCCTGGCCGCCCCCCTGGTGCTGCCCGA 19479

Qy      353 CGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGA 412
Db    19480 GCGCGGCGGCGTCCAACTCCAGGTCCGCGTCGGCGCCCCCGACGCCGCCGGCCGCCGCAC 19539

Qy      413 CTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCG 472
Db    19540 CCTCGGCATCTTCTCCCGCGTCGAGGACGGCTTCGACCTGCCCTGGTCGCAACACGCCAC 19599

Qy      473 CCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCT 532
Db    19600 CGGCGTCCTGACCGCCGGCGCCGGCGCCCCCGACCCCACCTTCGACGCCACCGTCTGGCC 19659

Qy      533 CTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGG 592
Db    19660 CCCCAGCGGCGCCGAACCCGTCGACCTCACCGGCGCGTACGAGCGCCTGGCCGCACTCGG 19719

Qy      593 GCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGC 652
Db    19720 CTTCCAGTACGGCCCCGCCTTCCAGGGCCTGCGCGCCGCCTGGCGCCGCGACACCGAGGT 19779

Qy      653 CATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCA 712
Db    19780 CTACGCCGAAGTGGCCCTGCCCGACGGCGCGGACACCGACCCCGCCGCCTTCGGACTGCA 19839

Qy      713 TCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGC 772
Db    19840 CCCGGCCCTGCTGGACGCCGCACAACACGCCGC---CGCCTACGCCGACCTCGGCGCCAT 19896

Qy      773 CCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGC 832
Db    19897 CAGCCGCGGCGGCCTGCCGTTCGCCTGGGAAGGCGTCTCGCTCGCCGCCGCCGGCGCCAC 19956

Qy      833 GCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCC 884
Db    19957 CACCGTCCGCGCCCGGATCGCCCCGGCCGGCGAGGACACCGTCACCATCGCC 20008
```

RESULT 10
US-10-937-379-48
; Sequence 48, Application US/10937379
; Publication No. US20050176653A1
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: McAlpine, James B.
; APPLICANT: Farnet, Chris M.
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Dan Sorensen
; TITLE OF INVENTION: POLYENE POLYKETIDES AND METHODS OF PRODUCTION
; FILE REFERENCE: 3010-4US
; CURRENT APPLICATION NUMBER: US/10/937,379
; CURRENT FILING DATE: 2004-09-10
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 5766
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
US-10-937-379-48

Query Match 11.5%; Score 114.6; DB 9; Length 5766;
Best Local Similarity 48.1%; Pred. No. 2e-18;
Matches 441; Conservative 0; Mismatches 449; Indels 27; Gaps 3;

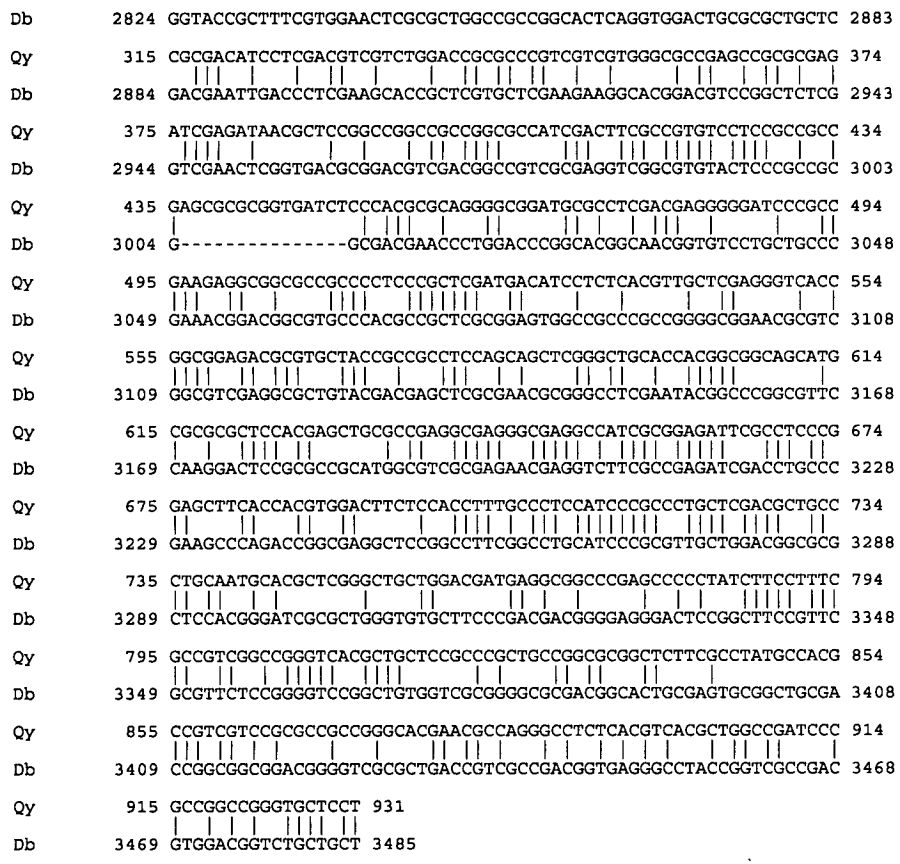
```
Qy       27 CGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCGCGAGCGTTACTGGTT- 85
Db     2584 CCGACGGCGGACTGGGTCGATCTGCCGACGTATCCGTTCCAGCGCAAGCGGTTCTGGCTG 2643

Qy       86 --------CTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAGGCGCTTG 137
Db     2644 GAAGCCGGGACCGCCGCCGGGGACGCGTCGGCGTTCGGGCAGACGGTGGTGGACCACCCG 2703

Qy      138 CACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGCCGGCTC 197
Db     2704 CTGCTCGGCGCCGTCGTCGCGGTCCCCGGGACCGGCGGGCTGCTGTACACCGGCCGGATC 2763

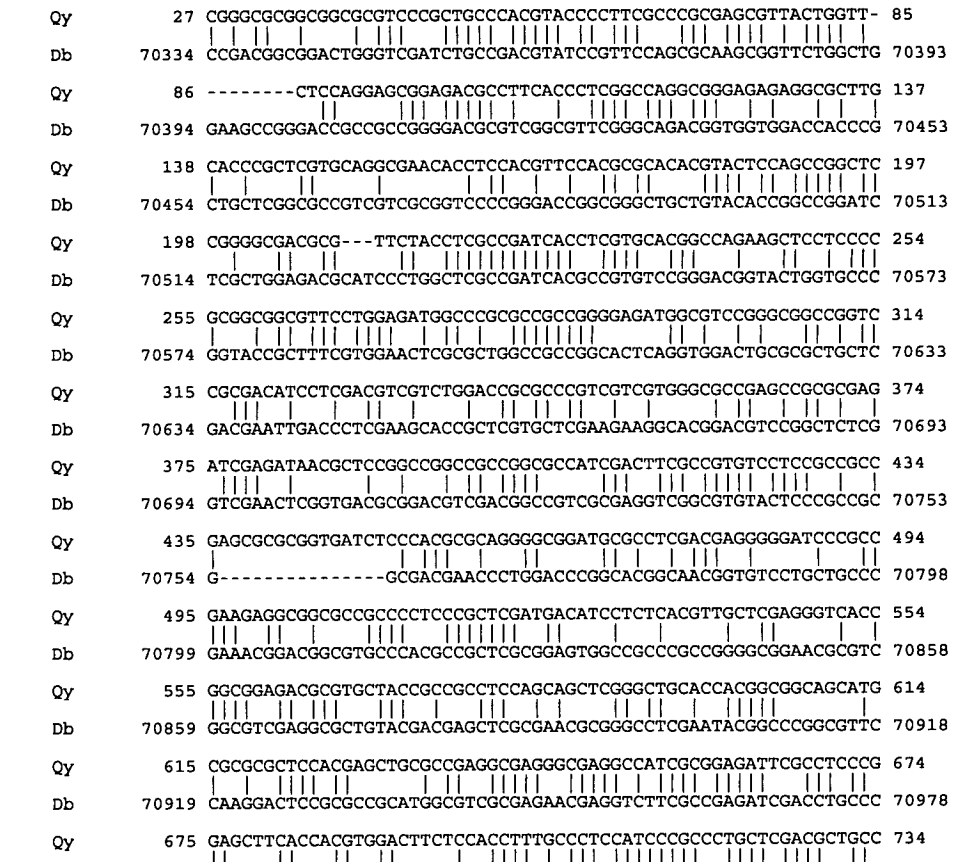
Qy      198 CGGGGCGACGCG---TTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCC 254
Db     2764 TCGCTGGAGACGCATCCCTGGCTCGCCGATCACGCCGTGTCCGGGACGGTACTGGTGCCC 2823

Qy      255 GCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTC 314
```



RESULT 11
US-10-937-379-24
; Sequence 24, Application US/10937379
; Publication No. US20050176653A1
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: McAlpine, James B.
; APPLICANT: Farnet, Chris M.
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Dan Sorensen
; TITLE OF INVENTION: POLYENE POLYKETIDES AND METHODS OF PRODUCTION
; FILE REFERENCE: 3010-4US
; CURRENT APPLICATION NUMBER: US/10/937,379
; CURRENT FILING DATE: 2004-09-10
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 73599
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
US-10-937-379-24

Query Match 11.5%; Score 114.6; DB 9; Length 73599;
Best Local Similarity 48.1%; Pred. No. 1.2e-18;
Matches 441; Conservative 0; Mismatches 449; Indels 27; Gaps 3;



RESULT 12
US-10-819-386A-1/c
; Sequence 1, Application US/10819386A
; Publication No. US20050089884A1
; GENERAL INFORMATION:
; APPLICANT: Korea advanced Institute of Science and Technology
; APPLICANT: Shanghai Jiaotong University
; APPLICANT: LEE, Sang Yup

; APPLICANT: DENG, ZIXIN
; APPLICANT: CHEN, SHI
; APPLICANT: JEONG, Ki Jun
; APPLICANT: ZHOU, XIUFEN
; TITLE OF INVENTION: Genes for the Synthesis of FR-008 Polyketides
; FILE REFERENCE: PO03-B015
; CURRENT APPLICATION NUMBER: US/10/819,386A
; CURRENT FILING DATE: 2004-04-06
; PRIOR APPLICATION NUMBER: KR10-2003-0074035
; PRIOR FILING DATE: 2003-10-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 138203
; TYPE: DNA
; ORGANISM: Streptomyces sp. FR-008
US-10-819-386A-1

Query Match 10.9%; Score 109.2; DB 9; Length 138203;
Best Local Similarity 50.0%; Pred. No. 2.2e-17;
Matches 384; Conservative 0; Mismatches 373; Indels 11; Gaps 4;

```
Qy        212 CTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGA 271
              ||  |||||||| |||    |||| ||||      ||| ||  |||   | || ||||| ||
Db     133269 CTGGCTCGCCGACCACACGGTGCGCGGCACCGTGCTGCTGGCCGGCACCGCCTTCCTCGA 133210

Qy        272 GATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGT 331
                | |||  |  |||||| ||   ||     | |    |  |||   ||  ||  |  ||
Db     133209 ACTCGCCGTCCGCGCCGGCGACGAGGTCGGCTGCGACCGCGTCGAGGAACTCACCCTCGC 133150

Qy        332 CGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCG 391
              || |  |  ||| |    ||  |  |||| ||  | |  | || ||  |||       ||
Db     133149 CGCCCCGCTCGCCCTGCCCGAGGACGGCGGCGTCCAGGTCCAGGTCTGGATCGGCAGCCC 133090

Qy        392 GCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTC 451
                |  |||||  |  || | |  || ||| ||||  | ||||  |||  |    |
Db     133089 CGACGAGACCGGCCGCCGCGCCCTGACCCTGTACTCCCGCCCCGACGGCGACGACGACCG 133030

Qy        452 CCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCC 511
              | |  |    ||       | |  |  |    |  |||| |  |  |  |  |  |
Db     133029 GCCCTGGACCCAGCACGCCGCCGGCACCCTGACCAACGGCGCCCACCACGGCGAGGAGTT 132970

Qy        512 CCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTA 571
              | | ||  |          | | |  ||  |  ||| ||   ||||   |||| | |||
Db     132969 CGACGCGGCGGTCTGGCCGCCCGCCGGTGCCGAGAGTGT---CGGCCTCGACGGCTTCTA 132913

Qy        572 CCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCT 631
              |    | ||    ||  |||  |   | |||||  |  | | ||||  || | ||   |
Db     132912 CGAGGGGCTGGCCGAGTCCGGCTTCGCCTACGGCCCCGCCTTCCGCGGCCTGCGCGCCGT 132853

Qy        632 GCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGA 691
                | ||  | |  |  ||| || |||| ||| |    |||  | |   | |   | ||
Db     132852 CTGGCGCCGGGGCGAGGAGACCTTCGCCGAGGTGGCTCTCGACGCGGGACCCGCGGGCGA 132793

Qy        692 CTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGG 751
              |  |||  | || | |||| || || |||| ||||| |||||||||    |||   |
Db     132792 CGCCTCTTCGTTCGGCCTGCACCCGGCGCTGCTGGACGCGGCCCTGCACGCCACCGCCCT 132733

Qy        752 GCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCAC 811
              | |  ||  ||||  |||||  | |  |   || || |||||          |||||
Db     132732 CGCCCCGCTCGGCGAGGACGCCCGCGGCGGC---CTGCCGTTCGCCTGGCAGGACGTCAC 132676

Qy        812 GCTGCTC-CGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGC 870
              || | | || |||  |||  || ||| |    ||     | ||||  | |  || |||
Db     132675 CCTCCACGCGGCCGGCGCCACCGAGGCCCGGGTCCGCATCACCCCCGCGGGCGACGACGC 132616

Qy        871 CGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCC 930
              ||      | |  ||  | | |  ||  |||   |||      ||||      |  ||
Db     132615 CG----TCACCCTCACCGTCGCGGACACCACCGGGGCACCCGTCGCCTCGGTCGGCTCCC 132560

Qy        931 TCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGC 978
              |||  | ||||      | |||  |   | |  | | || || ||  ||
Db     132559 TCGTGCTGCGCTCCGCCCCCGACGCGCGCCGGGCCGCCGCCGGCGTGC 132512
```

RESULT 13
US-10-760-493-30
; Sequence 30, Application US/10760493
; Publication No. US20050187167A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia BioSciences Inc
; APPLICANT: Farnet, Chris M.
; APPLICANT: McAlpine, James B.
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Bachmann, Brian O.
; APPLICANT: Piraee, Mahmood
; TITLE OF INVENTION: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
; TITLE OF INVENTION: PHARMACEUTICALS
; FILE REFERENCE: 3004-9US
; CURRENT APPLICATION NUMBER: US/10/760,493
; CURRENT FILING DATE: 2004-01-21
; PRIOR APPLICATION NUMBER: USSN 60/441,123
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: USSN 60/469,810
; PRIOR FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: USSN 60/491,516
; PRIOR FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: USSN 60/494,568
; PRIOR FILING DATE: 2003-08-13
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 15624
; TYPE: DNA
; ORGANISM: Streptomyces aizunensis
US-10-760-493-30

Query Match 10.5%; Score 105.2; DB 9; Length 15624;
Best Local Similarity 46.8%; Pred. No. 3e-16;
Matches 412; Conservative 0; Mismatches 453; Indels 15; Gaps 2;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | |||||| || || ||  |||  |    |    |||||||||   || |||||  |||
Db       7334 TCGACTGGCAGGCGTTCTTCGCCGGGACCGGCGCCCGGCGCGTCGACCTCCCCACCTACG 7393

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
              |||||   ||| |||| ||||||  | ||      |     |  | ||  |  || | |||
Db       7394 CCTTCCAGCGCCAGCGCTACTGGCCCGCCGTCTCCTCCCTCTACCTCGGCGACGTCGAGG 7453

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCT----------CCACGT 168
              ||  | |    |    | ||| |  || |   |   |                ||  ||
Db       7454 CGATCGGGCTCGACGACACCGCGCACCCGCTGCTCAGTGCGGGTGTCGCCCTGCCCGAGT 7513

Qy        169 TCCACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGC---GTTCTACCTCGCCGATC 225
              | |||       || | || | ||||||  |  |  | |       ||  |||||||  |
Db       7514 CCGACGGCATGGTGTTCGCCGGGCGGCTCGCGCTCTCCACCCACGCCTGGCTCGCCGACC 7573

Qy        226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285
              ||  | | | ||||      || || ||||   |||| ||| | ||| |||||  |  ||
Db       7574 ACGCCATCCTCGGCAGCGTCCTGCTGCCCGGTACGGCCTTCGTCGAGCTGGCCACCCGCG 7633

Qy        286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345
              |||| ||   ||  | |      |  ||  |  ||  |    ||    ||  |
Db       7634 CCGGCGACCAGGTCGGCTGCGATTACCTGGAAGAGCTGACCCTCGAAGCGCCCCTCGTCC 7693

Qy        346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405
                ||     |||| ||  | ||      | | |     |      |  |||||
Db       7694 TGCCCGAGCACGGCGGCGTCCAGCTGCGCGTGTGGGTCGGCGCCGCCGACGAGTCCGGCC 7753

Qy        406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465
              |||||| ||  ||||  |||||  || |  | ||  |    |  |  ||  |  ||
Db       7754 GACGGCCGTTCGCCCTGCACTCCCGGGCCGAAGGCCTGCCGGTCGAGGAGCCGTGGACGC 7813

Qy        466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525
              ||     | ||  |     | ||  |  || |  || |  | ||  |    ||    |
Db       7814 GGCACGCCGGCGGTGTACTCGCCGAAGGCGGGCGGCCCCCGGCCGACTTCGACCTGACGG 7873
```

```
Qy        526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585
Db       7874 CCTGGCCCCCGCCGGGCGCCGTCGAAGTGGACCTTGACGGGCGCTACGACCAGCTCGACG 7933

Qy        586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645
Db       7934 GCATCGGCTTCGCCTATGGCCCCACCTTCCGTGGCCTGCGTACGGCCTGGCAGCTCGACG 7993

Qy        646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705
Db       7994 GCGAGATCTACGCCGAGGTCAGGCTGCCCGAGGGAGCCGAGGGCGAGGCGGGCCGGTTCG 8053

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
Db       8054 GCCTGCACCCGGCCCTGCTCGACGCGGCACTGCACGCCATCGGGCTGGGCGGCCTCGGCG 8113

Qy        766 AGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGC 825
Db       8114 CCGACGACGGCCAGGGGAGGCTCCCCTTCGCCTGGAGCGGAGTATCGCTGCACGCGGGCG 8173

Qy        826 TGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGC 865
Db       8174 GGGCTGCCGCACTGCGCGTCCACCTCGCTCCGGCGGGCGC 8213


RESULT 14
US-10-760-493-32
; Sequence 32, Application US/10760493
; Publication No. US20050187167A1
; GENERAL INFORMATION:
;  APPLICANT: Ecopia BioSciences Inc
;  APPLICANT:  Farnet, Chris M.
;  APPLICANT:  McAlpine, James B.
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Bachmann, Brian O.
;  APPLICANT:  Piraee, Mahmood
;  TITLE OF INVENTION: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
;  TITLE OF INVENTION:  PHARMACEUTICALS
;  FILE REFERENCE: 3004-9US
;  CURRENT APPLICATION NUMBER: US/10/760,493
;  CURRENT FILING DATE:  2004-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/441,123
;  PRIOR FILING DATE: 2003-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/469,810
;  PRIOR FILING DATE: 2003-05-13
;  PRIOR APPLICATION NUMBER: USSN 60/491,516
;  PRIOR FILING DATE: 2003-08-01
;  PRIOR APPLICATION NUMBER: USSN 60/494,568
;  PRIOR FILING DATE: 2003-08-13
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
;   LENGTH: 16299
;   TYPE: DNA
;   ORGANISM: Streptomyces aizunensis
US-10-760-493-32

  Query Match             10.5%;  Score 105;  DB 9;  Length 16299;
  Best Local Similarity   46.5%;  Pred. No. 3.4e-16;
  Matches  477;  Conservative    0;  Mismatches  515;  Indels   33;  Gaps    3;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
Db       2657 TCGACTGGGACGCGTACTACTCCGGTACGGACGTCCAGCGCGTCGACCTGCCCACCTACG 2716

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAG---------CGGAGACGCCTTCACCC 111
Db       2717 CCTTCCAGCGCGCGCACTACTGGCTCGACGCAGGCCGGCCCCTCGGCGACGTCTCCTCGG 2776

Qy        112 TCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCC 171
Db       2777 CCGGGCTCGGTGCGGCCGGCCACCCGCTGCTCGGGGCCGCCGTGGCCCTCGCCGACCTCG 2836

Qy        172 ACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTAC---CTCGCCGATCACC 228
Db       2837 ACGGTTTCCTCTACACCGGCCGTCTCTCGCTCGACACCCACCCCTGGCTCGCCGACCACG 2896

Qy        229 TCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCG 288
Db       2897 CCGTCATGGGTTCGGCCGTACTGCCGGGCACCGCCTTCGTCGAACTGGCCATCCGCGCCG 2956

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
Db       2957 GTGACCAGGTCGGCTGCGACCTGCTCGAAGAACTCACCCTGCACGCACCGCTCGTACTGC 3016

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408
Db       3017 CCCCGGCCGGAGGTGTGCAGGTCCAGTTGTGGGTCGGCGCACCGGACGCCACCGGCCGCC 3076

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCAC------------ 455
Db       3077 GCACCCTGGGTGTGCACTCCCGCCCCGAGCCCGCACCGGACGCCGTCGGCCCGGACGCCG 3136

Qy        456 --------GCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGC 507
Db       3137 ACGCGGCGGAGCCGTGGACCCGGCACGCCGACGGTGTGCTCGCCACGGGTGCCCCGCAGC 3196

Qy        508 CGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGT 567
Db       3197 CGTCCTTCGCCCCCGACGTCTGGCCGCCGGCCGGTGCCAGGCCCCTGCCCGTCGACGAGC 3256

Qy        568 GCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACG 627
Db       3257 TGTACGCCGGGCTCGCCGAGGCGGGCCTCGAATACGGCCCCGCCTTCCAGGGCGTCCGCG 3316

Qy        628 AGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACG 687
Db       3317 CGGCCTGGGCGAGCGACGACGCGGCCTACGTCGAGATCGCGGCCGCCGACGGACAGTGGG 3376

Qy        688 TGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGC 747
Db       3377 CCGATGCCCCGCTGTTCGGACTGCATCCCGCGCTCCTCGACTCGGCGCTGCACGCCATCG 3436

Qy        748 TCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGG 807
Db       3437 GTCTGGCCGGGCTCGTCGAGGACACCGGCCGCGGCCGGCTGCCCTTCTCCTGGTCCGGGG 3496

Qy        808 TCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGC 867
Db       3497 TGTCCCTGTACGCCGTGGGCGCCTCGGTGCTGCGCGTACGGCTGGCCAAGGCCGGACCGG 3556

Qy        868 CGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGC 927
Db       3557 ACGCGGTGTCCCTGGCCCTCGCCGACGGCGCCGGACAGCCCGTGGGCGACATCGCCTCGC 3616

Qy        928 TCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCG 987
Db       3617 TCACCCTGCGCCCTGTCTCGGCCGAGCAGCTGGACACCGGGCGGGGCGGTCACCATGACG 3676

Qy        988 CCCAG 992
Db       3677 CGCTG 3681


RESULT 15
US-10-378-083-1
; Sequence 1, Application US/10378083
; Publication No. US20040053274A1
; GENERAL INFORMATION:
;  APPLICANT: President of Tokyo Institute of Technology
;  TITLE OF INVENTION: A gene cluster of vicenistatin biosynthase, a vicenisamine
;  TITLE OF INVENTION:  glycosyltransferase polypeptide, and a gene encoding the
;  TITLE OF INVENTION:  polypeptide
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/378,083
```

```
;  CURRENT FILING DATE:  2003-03-04
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 64492
;   TYPE: DNA
;   ORGANISM: Streptomyces halstedii sp.HC-34
US-10-378-083-1

  Query Match            10.4%;  Score 103.8;  DB 7;  Length 64492;
  Best Local Similarity   47.0%;  Pred. No. 5.1e-16;
  Matches  370;  Conservative    0;  Mismatches  402;  Indels   15;  Gaps    1;

Qy        212 CTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGA 271
              ||   | || |||||||  ||   |||   || ||| || || |   |||  ||| ||||
Db      50197 CTGGTTGGCGGATCACCAGGTCGCCGGGGTGACGCTGCTGCCGGGTACGGGCTTCGTGGA 50256

Qy        272 GATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGT 331
              | ||||   |   ||||| ||   ||    | |  | | | |    ||  |       |
Db      50257 GCTGGCGGTCCGTGCCGGTGACGAGGTCGGCTGTGGCCGGTTGGAGGAACTGACGCTGGA 50316

Qy        332 CGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCG 391
              |    |  ||       || |  |||||||  | || | || |||   |       |
Db      50317 GGCACCGCTCGTCGTGCCCGACCGGGGCGCCGTCCAGCTCCAGGTCGTCGTCGGCGCACT 50376

Qy        392 GCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTC 451
              |  ||   ||| || |  |  | ||  ||| | ||||  ||||||   |  |   |||
Db      50377 GGAGGAGTCCGACGTCCGCACCGTCAGCGTCTACTCCCGCGCCGAGGAGGACGACGTCT- 50435

Qy        452 CCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCC 511
                        |||  |||| || |   ||      || ||  |  |  |
Db      50436 --------------GGACCCGCCACGCCACCGGATTCGTCGGCGCCGCCGTCGACGGCGG 50481

Qy        512 CCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTA 571
              ||    || || |      | | | |   || |  |||     | ||    ||    |||
Db      50482 ACTGAACCTGGAGGCATGGCCCCCGCCGGGCGCCCGGGCGGTGGACGTCAGCGACGTCTA 50541

Qy        572 CCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCT 631
              |  | | ||     | |  ||        ||||   |  | | |  |  ||||    | |
Db      50542 CGCCGGACTGGCCGACCAGGGCTACGGATACGGTCCCGTCTTCCAGGGCCTCCGGTCGGT 50601

Qy        632 GCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGA 691
              | | || |||||  |||||||   |||| ||| |     | |||||   |   || | |||
Db      50602 GTGGCGCGGCGACGGCGAGGTGTTCGCCGAGGTCGTGTTGCCGGAAGGTGCACAGGCGGA 50661

Qy        692 CTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGG 751
              |  |||| | || |  || || ||||||||||| ||||| || |||||   |||   |
Db      50662 CGCCTCCGCATTCGGGCTGCACCCCGCCCTGCTGGACGCCGCACTGCACGCCACCGACTA 50721

Qy        752 GCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCAC 811
              ||       ||   || | |   | | |  |||| || ||||||     | | ||| |
Db      50722 CCTCGACCCGGAGTCCGCCGTCGAGGGCACGCATCTGCCGTTCGCCTGGACCGGTGTCTC 50781

Qy        812 GCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCC 871
              ||||| |  | |     || |    ||    |       |||  |  |   |
Db      50782 GCTGCACGCCACCGGTGCGTCCGAACTGCGGGTACGGATCACCGCCACCGGCAACGACGG 50841

Qy        872 GGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCT 931
                | ||  || |   |||      |  |   |   |||   || ||| || |  |||| |
Db      50842 ATACGCGCTCGACCTCGCCGACACCACCGGCCGCCCCGTCGCCACCGTCCAGTCGCTCGT 50901

Qy        932 CGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCA 991
              |   |  |    |  |  | |||   |  ||  | | ||    |  ||||| |
Db      50902 GCTGCGCCCGGTCACCGAGGAGCAGCTCCGCAGCGCCCAGGGCGGCAACGGCCCCGACAG 50961

Qy        992 GCGGCTC 998
                | | ||
Db      50962 CCTGTTC 50968


Search completed: April 27, 2006, 14:39:44
Job time : 1008.57 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 627.947 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_35000_35999
Perfect score: 1000
Sequence: 1 tggactgggagccgctcgac.....acccccgcccagcggctcta 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	110.6	11.1	113193	7	US-10-478-943D-1	Sequence 1, Appli
2	86.6	8.7	6360	14	US-11-056-470-2	Sequence 2, Appli
3	85.2	8.5	14055	7	US-10-496-351-1	Sequence 1, Appli
c 4	85.2	8.5	82746	7	US-10-496-351-56	Sequence 56, Appl
5	81.4	8.1	984	11	US-11-228-659-45	Sequence 45, Appl
6	81.4	8.1	4470	11	US-11-228-659-44	Sequence 44, Appl
7	75.8	7.6	10968	14	US-11-075-185-35	Sequence 35, Appl
8	75.8	7.6	78869	14	US-11-075-185-1	Sequence 1, Appli
c 9	74.2	7.4	14055	7	US-10-496-351-1	Sequence 1, Appli
10	74.2	7.4	82746	7	US-10-496-351-56	Sequence 56, Appl
11	72.2	7.2	2742	7	US-10-496-351-22	Sequence 22, Appl
12	71.2	7.1	2489	11	US-11-096-568A-14645	Sequence 14645, A
13	71.2	7.1	88421	14	US-11-205-109-1	Sequence 1, Appli
14	71	7.1	1192	11	US-11-096-568A-27171	Sequence 27171, A
15	70.4	7.0	5679	14	US-11-075-185-36	Sequence 36, Appl
16	70	7.0	37507	8	US-10-522-037-2	Sequence 2, Appli
17	69.8	7.0	9975	7	US-10-496-351-4	Sequence 4, Appli
c 18	69	6.9	4983	14	US-11-052-554A-521	Sequence 521, App
19	68.4	6.8	4770	7	US-10-496-351-3	Sequence 3, Appli
20	67.2	6.7	3354	7	US-10-496-351-12	Sequence 12, Appl
c 21	67	6.7	88421	14	US-11-205-109-1	Sequence 1, Appli
22	66.2	6.6	1734	11	US-11-096-568A-12766	Sequence 12766, A
23	66.2	6.6	116856	14	US-11-143-980-1	Sequence 1, Appli
24	66	6.6	9222	7	US-10-496-351-2	Sequence 2, Appli
25	65.4	6.5	1546	11	US-11-096-568A-23918	Sequence 23918, A
c 26	65.2	6.5	2742	7	US-10-496-351-22	Sequence 22, Appl
27	65	6.5	1014	11	US-11-228-659-10	Sequence 10, Appl
28	65	6.5	4595	11	US-11-228-659-8	Sequence 8, Appli
c 29	64.6	6.5	9222	7	US-10-496-351-2	Sequence 2, Appli
30	64	6.4	897	8	US-10-858-730-45	Sequence 45, Appl
31	63.4	6.3	975	11	US-11-228-659-48	Sequence 48, Appl
32	63.4	6.3	4114	11	US-11-228-659-47	Sequence 47, Appl
33	63.4	6.3	11070	14	US-11-075-185-34	Sequence 34, Appl
34	63.2	6.3	1245	11	US-11-096-568A-19250	Sequence 19250, A
35	63	6.3	1422	11	US-11-096-568A-12617	Sequence 12617, A
36	62.2	6.2	1590	8	US-10-858-730-162	Sequence 162, App
37	62.2	6.2	2247	7	US-10-496-351-8	Sequence 8, Appli
c 38	62.2	6.2	4770	7	US-10-496-351-3	Sequence 3, Appli
39	62	6.2	11905	7	US-10-496-351-57	Sequence 57, Appl
40	61.4	6.1	1614	11	US-11-096-568A-16842	Sequence 16842, A
41	61	6.1	1707	11	US-11-096-568A-18220	Sequence 18220, A
42	60.6	6.1	1029	7	US-10-496-351-51	Sequence 51, Appl
43	60.4	6.0	1566	14	US-11-143-980-7	Sequence 7, Appli
c 44	60.4	6.0	116856	14	US-11-143-980-1	Sequence 1, Appli
45	60.2	6.0	1485	14	US-11-143-980-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-10-478-943D-1
; Sequence 1, Application US/10478943D
; Publication No. US20060073574A1
; GENERAL INFORMATION:
; APPLICANT: University College Dublin, National University of Ireland
; TITLE OF INVENTION: Engineered biosynthesis of novel polyenes
; FILE REFERENCE: PC-1636US
; CURRENT APPLICATION NUMBER: US/10/478,943D
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: S2001/0527
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 113193
; TYPE: DNA
; ORGANISM: Streptomyces nodosus
US-10-478-943D-1

Query Match 11.1%; Score 110.6; DB 7; Length 113193;
Best Local Similarity 50.3%; Pred. No. 1.4e-14;
Matches 390; Conservative 0; Mismatches 364; Indels 21; Gaps 4;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | |||||||   || || |||   ||   |    ||   || |   |||||||| |||
Db      36261 TCGACTGGGCCGCGTTCCACGACGGCACCGGTGCCCGCACCGGCGAACTGCCCACCTACG 36320

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGA---------GACGCCTTCACCC 111
              |||||   |||||||||||||  |    || ||           ||||||  |   |
Db      36321 CCTTCCAGCACGAGCGTTACTGGCCCAAGGCGACCGCGACCGCTGTCGACGCCACCGGTC 36380

Qy        112 TCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCC 171
              ||||||  ||   |   |   ||  | |  ||   | |      ||  | |
Db      36381 TCGGCCTCGCCTCCGCCGACCATCCGCTGCTGGGCGCCGCGATGTCGGTGGCCGGTTCCG 36440

Qy        172 ACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGC---GTTCTACCTCGCCGATCACC 228
              |||  | |  |  | || ||   ||   |  || | |      ||  |||||||| |||
Db      36441 ACGAACTCCTGCTCACCGGCACCCTGTCGCTCGCCACCCACCCCTGGCTCGCCGACCACA 36500

Qy        229 TCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCG 288
```

```
Db      36501 GCGTCGACGGCATGGTCGTCTTCCCCGGTACCGGGTTCCTGGAGCTGGCGGTGCGCGCCG 36560

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348

Db      36561 CCGACCAGGCCGGCTGCGACCGGGTCCAGGAACTGACCGTCGCCACCCCGCTCGTGCTGC 36620

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCA 408

Db      36621 CCGCCACGGGCGCCGTGC------AGATGCAGATCTCCGTCGGCGCCGCCGGCGAGGACG 36674

Qy        409 TCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGA 468

Db      36675 GCTCCCGTGAGCTGCGCTTCTTCACCCGGCCCGGTGAGGACTTCGACGCCGAGTGGACCC 36734

Qy        469 TGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACA 528

Db      36735 AGCACGCCACCGGCCGCATCGCCTCCGGCGAGCACGTCCTCGGCTTCGACACCACGGTGT 36794

Qy        529 TCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGC 588

Db      36795 GGCCGCCGCGCGACTCCGAGGCCGTCGACATCGAGGAACTGTTCGACCGCTTCGCCTCCG 36854

Qy        589 TCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCG 648

Db      36855 ACGGTCTGGAGTACGGCCCGGTCTTCCGCGGACTGCGCGCCGCCTGGCGCCAGGACGACA 36914

Qy        649 AGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCC 708

Db      36915 CGGTCTACGCCGAGGTCGAACTGCCCGACTC---CGTCGAGGACGCCGGTGCCTTCGGTC 36971

Qy        709 TCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGA 763

Db      36972 TGCACCCGGCCCTGCTGCACGCCGCCCTGCACGGCACCGCCTTCCTGTCCGACGA 37026


RESULT 2
US-11-056-470-2
; Sequence 2, Application US/11056470
; Publication No. US20050261263A1
; GENERAL INFORMATION:
;  APPLICANT: SANTI, Daniel V.
;  APPLICANT:  MYLES, David C.
;  APPLICANT:  TIAN, Zong-Qiang
;  APPLICANT:  HUTCHINSON, C. Richard
;  APPLICANT:  JOHNSON, Robert G., Jr.
;  APPLICANT:  ZHOU, Yi-Qing
;  APPLICANT:  FENG, Li
;  TITLE OF INVENTION: BENZOQUINONE ANSAMYCINS
;  FILE REFERENCE: 300622007510
;  CURRENT APPLICATION NUMBER: US/11/056,470
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: US 10/212,962
;  PRIOR FILING DATE: 2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/310,779
;  PRIOR FILING DATE: 2001-08-06
;  PRIOR APPLICATION NUMBER: US 60/389,255
;  PRIOR FILING DATE: 2002-06-14
;  PRIOR APPLICATION NUMBER: US 60/393,929
;  PRIOR FILING DATE: 2002-07-03
;  PRIOR APPLICATION NUMBER: US 60/395,275
;  PRIOR FILING DATE: 2002-07-12
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 6360
;   TYPE: DNA
;   ORGANISM: Streptomyces geldanus
US-11-056-470-2

  Query Match             8.7%;  Score 86.6;  DB 14;  Length 6360;
  Best Local Similarity   46.9%;  Pred. No. 2.3e-09;
  Matches  424;  Conservative    0;  Mismatches  459;  Indels   22;  Gaps    4;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60

Db       2720 TGGACTGGGACGCGATGTTCGCCGGGTCCGGCGCCCGGCCCGTCGCCCTTCCCACGTACG 2779

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTT---------------CTCCAGGAGCGGAGACGCCT 105

Db       2780 CCTTCCAGCACGAGCACTACTGGCTGGAGCGGGCCGCCGGCTCCGGCGACGTGGGCGCGG 2839

Qy         106 TCACCCTCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCA 165

Db       2840 TGGGGCTCGGCGAGGCGGGCCATCCGCTGCTGGGCGCGGTGGTGCAGCT---CCCGGAGA 2896

Qy         166 CGTTCCACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATC 225

Db       2897 CGGGCGGGGTGCAGCTCAGCGGGCGGCTGTCGGTACGGGCCCAGCCCTGGCTGGGCGAAC 2956

Qy         226 ACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCG 285

Db       2957 ACGTCATCTCCGGGGCGGTGCTGGTGCCCGGCACCGCCATGGTGGAACTGGCCGTCCGCG 3016

Qy         286 CCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGC 345

Db       3017 CCGGGGACGAGACCGGCACCCCGGTGCTGGAGGAGCTGGTGATCGGGCAGCCGATGGTGC 3076

Qy         346 CCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCG 405

Db       3077 TGCCCGGCGACACCGCCCTCAGTGTCCAGGTCGTCGTGGGCGCGGACGAGGGCGGGCGGC 3136

Qy         406 CCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGC 465

Db       3137 GTACGGTGCGGATCTACTCCCGTACCGACGGGGGCACCGACTGGACCGAGCACGCCACCG 3196

Qy         466 GGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATG 525

Db       3197 GCACGCTCGCGGCGCAGGGCCCGGCACCGCTGGACGGGGCCGCGGGCGGGGCCGCCGTCG 3256

Qy         526 ACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGC 585

Db       3257 AGTGGCCGCCCGCGGAAGCCGAGCCGATCCCCGTGGAGGACTTCTACCGCTCGCTCGTCG 3316

Qy         586 AGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGG 645

Db       3317 ACGCCGGATACGCGTACGGACCGGCGTTCCGCGGGCTCGTCGCCGCGTGGCGCCGGGACG 3376

Qy         646 GCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTG 705

Db       3377 GTGAGATCTTCGGCGATGTGGCGCTGCCGGAGGCGTCCGTCGCGGAGGCCGAGCGGTTCG 3436

Qy         706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAAT---GCACGCTCGGGCTGCTGGACG 762

Db       3437 GCATCCACCCGGCGCTGCTGGACGCCGCACTGCACGCGGGCAGCTTCTGTCTGCCCTCCG 3496

Qy         763 ATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTC-CGC 821

Db       3497 ACCCGGCGCGACAGGTGACCCTGCTGCCGTTCGCCTGGAACACCGTGCGTCTGCACGCGG 3556

Qy         822 CCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAAC 881

Db       3557 GCGGCGCGTCCGCGGTCCGGGTGCATGTCCGCCCGGTCGGCGACGACGCCTTCTCGGTAC 3616

Qy         882 GCCAG 886

Db       3617 GCCTG 3621


RESULT 3
US-10-496-351-1
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
```

```
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             8.5%;  Score 85.2;  DB 7;  Length 14055;
  Best Local Similarity   45.0%;  Pred. No. 3.9e-09;
  Matches  442;  Conservative    0;  Mismatches  533;  Indels    7;  Gaps    3;

Qy          5 CTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTT 64
              ||  |  |||||  |    | ||||||    | ||  |  |   ||||  ||  | ||  ||||
Db       8443 CTCGACGCCGCGCTCGCCTCGCCGGAACCCGCCGTCCTCGCCGCCGCGCTCGACCCCCGC 8502

Qy         65 CGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGG 124
              |||  |    | |  |  |    ||  | |             |   |   |  |||  |  | ||  |
Db       8503 CGCGCCGCCGAGGTCCCGCCGCTGCTGCGCGACCTCGCCGGACCGCGCCGGGCGGTCGCG 8562

Qy        125 AGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTA 184
                 ||||   |  ||    |||  |   |  | ||   |  |   |  |  ||   ||||||||
Db       8563 GCCCAGGCCGTCGCCGGCGC-GGACCGGCCGCTGATCGACGTGGTCAGGGCGCACACCGC 8621

Qy        185 CTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAA 244
               ||     |    |   |||||||  |  |   |  |  |  ||   |||  ||  |   |     ||
Db       8622 GACCGTGCTCGGGCACCGCGACGCCTCCGCCGTGGACGCGAACCGCGCGTTCAAGGAGCT 8681

Qy        245 GCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGG 304
              |  |  ||  | ||     ||  | |||||  ||  |  |      ||  |     | |
Db       8682 GGGCGTCGACTCGCTCACGGCCGTGGAGCTGCGCAACCGGCTCGCCGCCGCGACCGGGCT 8741

Qy        305 GCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGA 364
               ||  | |  |  |     |   || |  |  ||||  ||  |  |  ||   |    |
Db       8742 CGCGCTCGCCGCCACCTCGGTGTTCGACCACCCGACCCCGGCGCTGCTCACCGCGCACCT 8801

Qy        365 GCCGCGCGAGATCGAGATAACGCTCCGGCCGG---CCGCCGGCGCCATCGACTTCGCCGT 421
               ||    ||||  |||  |    | |  |     |  ||  |||   ||  |  |  |||||
Db       8802 GCTCGCCGAGCTCGGGCACGCCCCCGCCGCCGGGACCACCGCGCCCCCCGCCACCGCCGC 8861

Qy        422 GTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGA 481
              |  ||||||  ||  |  |  |  |   |||  |   ||  |   |  |     |||  ||  ||
Db       8862 CGTCACCGCCGACGCGGGGGAGCCCATCGCGATCGTGTCGATGGCCTGCCGCTTCCCCGG 8921

Qy        482 GGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTG 541
              |   |   ||  ||   ||||    |  |  ||||      |  |  |     |  |  |
Db       8922 CGACGTGTCCTCCCCGGAGGACCTGTGGGACCTCGTCGCGGGCGGCGGGGACGCGCTCAC 8981

Qy        542 CTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCA 601
              | ||      | |||  |  |  |  |  |  |          |||        |  ||  |  |  |||
Db       8982 CCCGTTCCCCGCCGACCGGGGCTGGGACCTGGCGGCCCTGTTCGCCCCCGACCCGGACCG 9041

Qy        602 CGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGA 661
              ||  |||  |||  |  |||  |  |  |  |    |  ||  |  |||  |  ||  ||  |||  |
Db       9042 CGCCGGGCGCAGCCACGCCCGCGAGGGCGGGTTCCTCGCCGACGTCGCGGGCTTCGACGC 9101

Qy        662 GATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCT 721
                 |   ||          | ||  ||  ||  |  ||  ||||    |  |  ||  |     ||
Db       9102 CGACCTGTTCGGCATCAGCCCCCGCGAGGCCATCGGCACCGACCCGCAGCAGCGGCTGCT 9161

Qy        722 GCTCGACGCTGC---CCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGC 778
              |||  ||  |  ||      ||  |  |  ||  ||  |     |  |  |    ||  ||   ||  |
Db       9162 GCTGGAGGTGGCGTGGGAGCTGTTCGAGCGCGCGGGCATCGCCCCCGACTCGCTGCGCGG 9221

Qy        779 CCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCT 838
              | ||  |         |   ||  |||     ||  |||     |  ||||  |||  |  |     |
Db       9222 CACCGACACCGGCGTGTTCGCGGGCGTCATCGCGCACGGCTACCCGGTGCTGCCCGCCCA 9281

Qy        839 CTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGT 898
              ||  ||  |  |  |     |  |  |  |  |  |  |  ||    |  ||  ||  |        ||
Db       9282 CTCCGACCAGGGGCTGGAGGGGCACCGGGTGACCGGGGTGTCCGGGAGCGTGGTGTCCGG 9341

Qy        899 CACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCT 958
              |  |  |  |||  |   ||  |||  |  |     |  |    |  |  ||||    |   |||||  |  ||
Db       9342 CCGGGTCGCCTACGCCTTCGGGCTGCAGGGACCGGCGATGACCGTGGACACCGCCTGCTC 9401

Qy        959 CGCGACGGCGGACGTCGCGCCC 980
                ||  ||   ||  |  ||  ||  |
Db       9402 GTCGTCGCTGGTGGGCGTGCAC 9423


RESULT 4
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
;   LENGTH: 82746
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

  Query Match             8.5%;  Score 85.2;  DB 7;  Length 82746;
  Best Local Similarity   45.0%;  Pred. No. 2.8e-09;
  Matches  442;  Conservative    0;  Mismatches  533;  Indels    7;  Gaps    3;

Qy          5 CTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTT 64
              ||  |  |||||  |    | ||||||    | ||  |  |   ||||  ||  | ||  ||||
Db      28599 CTCGACGCCGCGCTCGCCTCGCCGGAACCCGCCGTCCTCGCCGCCGCGCTCGACCCCCGC 28540

Qy         65 CGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGG 124
              |||  |    | |  |  |    ||  | |             |   |   |  |||  |  | ||  |
Db      28539 CGCGCCGCCGAGGTCCCGCCGCTGCTGCGCGACCTCGCCGGACCGCGCCGGGCGGTCGCG 28480

Qy        125 AGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTA 184
                 ||||   |  ||    |||  |   |  | ||   |  |   |  |  ||   ||||||||
Db      28479 GCCCAGGCCGTCGCCGGCGC-GGACCGGCCGCTGATCGACGTGGTCAGGGCGCACACCGC 28421

Qy        185 CTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAA 244
               ||     |    |   |||||||  |  |   |  |  |  ||   |||  ||  |   |     ||
Db      28420 GACCGTGCTCGGGCACCGCGACGCCTCCGCCGTGGACGCGAACCGCGCGTTCAAGGAGCT 28361

Qy        245 GCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGG 304
              |  |  ||  | ||     ||  | |||||  ||  |  |      ||  |     | |
Db      28360 GGGCGTCGACTCGCTCACGGCCGTGGAGCTGCGCAACCGGCTCGCCGCCGCGACCGGGCT 28301

Qy        305 GCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGA 364
               ||  | |  |  |     |   || |  |  ||||  ||  |  |  ||   |    |
Db      28300 CGCGCTCGCCGCCACCTCGGTGTTCGACCACCCGACCCCGGCGCTGCTCACCGCGCACCT 28241

Qy        365 GCCGCGCGAGATCGAGATAACGCTCCGGCCGG---CCGCCGGCGCCATCGACTTCGCCGT 421
               ||    ||||  |||  |    | |  |     |  ||  |||   ||  |  |  |||||
Db      28240 GCTCGCCGAGCTCGGGCACGCCCCCGCCGCCGGGACCACCGCGCCCCCCGCCACCGCCGC 28181

Qy        422 GTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGA 481
              |  ||||||  ||  |  |  |  |   |||  |   ||  |   |  |     |||  ||  ||
Db      28180 CGTCACCGCCGACGCGGGGGAGCCCATCGCGATCGTGTCGATGGCCTGCCGCTTCCCCGG 28121

Qy        482 GGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTG 541
              |   |   ||  ||   ||||    |  |  ||||      |  |  |     |  |  |
```

```
Db      28120 CGACGTGTCCTCCCCGGAGGACCTGTGGGACCTCGTCGCGGGCGGCGGGGACGCGCTCAC 28061

Qy        542 CTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCA 601
              | ||     | ||| | | | |   | |          |||        | ||  | | |||
Db      28060 CCCGTTCCCCGCCGACCGGGGCTGGGACCTGGCGGCCCTGTTCGCCCCCGACCCGGACCG 28001

Qy        602 CGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGA 661
              || |||  |||  | ||| | | | |     |  ||  | ||| | || || | |||  |
Db      28000 CGCCGGGCGCAGCCACGCCCGCGAGGGCGGGTTCCTCGCCGACGTCGCGGGCTTCGACGC 27941

Qy        662 GATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCT 721
                  |   ||          | || || || | ||  ||||    | |  || |     ||
Db      27940 CGACCTGTTCGGCATCAGCCCCCGCGAGGCCATCGGCACCGACCCGCAGCAGCGGCTGCT 27881

Qy        722 GCTCGACGCTGC---CCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGC 778
              ||| || |  ||      ||  | |  || || |    | | |   ||  ||   || |
Db      27880 GCTGGAGGTGGCGTGGGAGCTGTTCGAGCGCGCGGGCATCGCCCCCGACTCGCTGCGCGG 27821

Qy        779 CCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCT 838
              | ||| |      |   ||  |||    || |||    |  |||| ||| | |    |
Db      27820 CACCGACACCGGCGTGTTCGCGGGCGTCATCGCGCACGGCTACCCGGTGCTGCCCGCCCA 27761

Qy        839 CTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGT 898
              || || | | |    |  | |  | | ||  |  ||  || |       ||
Db      27760 CTCCGACCAGGGGCTGGAGGGGCACCGGGTGACCGGGGTGTCCGGGAGCGTGGTGTCCGG 27701

Qy        899 CACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCT 958
              |  | | ||| |  ||  ||| | | |   |  ||||   |    ||||| | ||
Db      27700 CCGGGTCGCCTACGCCTTCGGGCTGCAGGGACCGGCGATGACCGTGGACACCGCCTGCTC 27641

Qy        959 CGCGACGGCGGACGTCGCGCCC 980
                || ||  ||  | || || |
Db      27640 GTCGTCGCTGGTGGGCGTGCAC 27619
```

RESULT 5
US-11-228-659-45
; Sequence 45, Application US/11228659
; Publication No. US20060064786A1
; GENERAL INFORMATION:
; APPLICANT: Brugiere, Norbert
; TITLE OF INVENTION: Isopentenyl Transferase Sequences and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 1507
; CURRENT APPLICATION NUMBER: US/11/228,659
; CURRENT FILING DATE: 2005-09-16
; PRIOR APPLICATION NUMBER: 60/610,656
; PRIOR FILING DATE: 2004-09-17
; PRIOR APPLICATION NUMBER: 60/637,230
; PRIOR FILING DATE: 2004-12-17
; PRIOR APPLICATION NUMBER: 60/696,405
; PRIOR FILING DATE: 2005-07-01
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: OsIPT2 coding sequence (018830_1)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(984)
US-11-228-659-45

Query Match 8.1%; Score 81.4; DB 11; Length 984;
Best Local Similarity 43.9%; Pred. No. 3.9e-08;
Matches 407; Conservative 0; Mismatches 511; Indels 9; Gaps 1;

```
Qy         63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
              ||  ||  | |||   |     ||| |  |  ||| | | ||  ||||  |        |
```

```
Db         40 TTGTCCACCAAGCCCAAGGTTATCTTCGTGCTCGGCGCCACCGCCACCGGCAAGTCCAAG 99

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
                | |  ||| |   | ||| | |  | |   |  |  ||  | | ||  |  | |
Db        100 CTCGCCATCCGCCTCGCCGCGCGCTTCGACGGCGAGGTCATCAACTCCGACAAGATCCAG 159

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
                | | |  ||| || | | |  | |  || |||| |           ||||
Db        160 GCGCACGACGGCTTCCCGGTCATCACCAACAAGGTCACCGACGAGGAGCGTGCCGGCGTC 219

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
              || || || |   |  |  ||          |||  |  | |||  ||||          |
Db        220 GCGCACCACCTCCTCGGCGGCGTCAGCCCCGACGCCGACTTCACCGCGGAGGACTTCCGC 279

Qy        303 GGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCC 362
                |  | ||| |  || | ||  |  ||||  |  | |||  | | ||  ||| | ||||
Db        280 CGCGAGGCGGCCGCCGCCGTCGCCCGCGTCCACGCGGCCGGCCGCCTCCCCGTCGTCGCC 339

Qy        363 GAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTG 422
              |    |    | ||| |   |    |    |||||  |||| |  ||  |||  || |
Db        340 GGCGGGTCGAACATCTACGTCGAGGCGCTCGTGGCCGGCGGCGGCGGCGCGTTCCTCGCG 399

Qy        423 TCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAG 482
              | | || |  |    |  | || |  |||||  ||  | | ||||      ||  |
Db        400 GCGTACGACTGCCTCTTCCTGTGGACCGACGTCGCGCCGGACCTGCTGCGGTGGTACACG 459

Qy        483 GGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGC 542
              | ||  | ||  || ||   || || ||  |  | | | |  ||   ||          ||
Db        460 GCGGCGCGCGTGGACGACATGGTGCGGCGCGGGCTGGTTGGCGAGGCCC---------GC 510

Qy        543 TCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCAC 602
              |  || ||  || ||| |  |  | |  |||| || | |  |      |  ||  | |
Db        511 GCCGGGTTCGACGCCGGGGCGGACTACACCCGCGGCGTGCGCCGCGCCATCGGGCTACCC 570

Qy        603 GGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAG 662
              |       |  |||    ||||       | ||||  ||||  |    ||  ||
Db        571 GAGATGCACGGCTACCTGCTGGCGGAGCGCGAGGGCGGCGCCGGCGCCGAGGACGACGAC 630

Qy        663 ATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTG 722
                | ||||  | |   | | | | | | | |  |    |      | |  |  ||
Db        631 GACCTCCTCGCCGGCATGCTCGAGGCCGCCGTGCGCGAGATCAAGGACAACACGTTCCGC 690

Qy        723 CTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCC 782
              |||  ||  | | |    || | || |||  ||   ||    | | |  | | |  |
Db        691 CTCACCGTGTCGCAGGTGGCCAAGATCCGGCGCCTCAGCGCGCTGCCCGGGTGGGACGTC 750

Qy        783 TATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTC 842
              |       |||    |   |||  |||  |      |    ||  | |   |
Db        751 CGGCGCGTGGACGCGACGGCGGTGGTGGCGCGCATGGCGGAGGGCGCGCCCCACGGCGAG 810

Qy        843 GCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACG 902
              | |       |  |  ||  | ||||   |    ||  | |||   ||  | ||  |||
Db        811 ACGTGGAGGGAGGTGGTGTGGGAGCCGTGCGAGGAGATGGTCAGCCGCTTCCTCGAGACG 870

Qy        903 CTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCG 962
              | ||||   ||||| |   ||   |  ||  |      ||  | | |||| ||  ||||
Db        871 CCCGCCGCCGCCGCCGCCGTCGTTGCCAACGGCAAAGTCGACGTGAACGTCGGCGACGCG 930

Qy        963 ACGGCGGACGTCGCGCCCACCCCCGCC 989
              | || | | |   |    |  |||||
Db        931 GCCGCCGGCTTGCCTGAGGCTGCCGCC 957
```

RESULT 6
US-11-228-659-44
; Sequence 44, Application US/11228659
; Publication No. US20060064786A1
; GENERAL INFORMATION:
; APPLICANT: Brugiere, Norbert
; TITLE OF INVENTION: Isopentenyl Transferase Sequences and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 1507
; CURRENT APPLICATION NUMBER: US/11/228,659

```
;  CURRENT FILING DATE:  2005-09-16
;  PRIOR APPLICATION NUMBER: 60/610,656
;  PRIOR FILING DATE: 2004-09-17
;  PRIOR APPLICATION NUMBER: 60/637,230
;  PRIOR FILING DATE: 2004-12-17
;  PRIOR APPLICATION NUMBER: 60/696,405
;  PRIOR FILING DATE: 2005-07-01
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 4470
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1484)...(2470)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: OsIPT2 full length sequence (018830_1)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2876, 2877, 2878, 2879, 2880, 2881, 2882, 2883, 2884, 2885,
;   LOCATION: 2886, 2887, 2888, 2889, 2890, 2891, 2892, 2893, 2894, 2895,
;   LOCATION: 2896, 2897, 2898, 2899, 2900, 2901, 2902, 2903, 2904, 2905,
;   LOCATION: 2906, 2907, 2908, 2909, 2910, 2911, 2912, 2913, 2914
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2915, 2916, 2917, 2918, 2919, 2920, 2921, 2922, 2923, 2924,
;   LOCATION: 2925, 2926, 2927, 2928, 2929, 2930, 2931, 2932, 2933, 2934,
;   LOCATION: 2935, 2936, 2937, 2938, 2939, 2940, 2941, 2942, 2943, 2944,
;   LOCATION: 2945, 2946, 2947, 2948, 2949, 2950, 2951, 2952, 2953
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2954, 2955, 2956, 2957, 2958, 2959, 2960, 2961, 2962, 2963,
;   LOCATION: 2964, 2965, 2966, 2967, 2968, 2969, 2970, 2971, 2972, 2973,
;   LOCATION: 2974, 2975
;   OTHER INFORMATION: n = A,T,C or G
US-11-228-659-44

  Query Match             8.1%;  Score 81.4;  DB 11;  Length 4470;
  Best Local Similarity   43.9%;  Pred. No. 3e-08;
  Matches  407;  Conservative    0;  Mismatches  511;  Indels    9;  Gaps    1;

Qy          63 TTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCG 122
               ||  ||  | |||   |     ||| |  |  ||| | | ||  ||||  |     |
Db        1523 TTGTCCACCAAGCCCAAGGTTATCTTCGTGCTCGGCGCCACCGCCACCGGCAAGTCCAAG 1582

Qy         123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
                 |  |||  |   | |||  |  |  |  |   |  ||  | | ||  |  |  |
Db        1583 CTCGCCATCCGCCTCGCCGCGCGCTTCGACGGCGAGGTCATCAACTCCGACAAGATCCAG 1642

Qy         183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
                   | | |  ||| ||  | |  |  | | || ||||  |          ||||
Db        1643 GCGCACGACGGCTTCCCGGTCATCACCAACAAGGTCACCGACGAGGAGCGTGCCGGCGTC 1702

Qy         243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
                 || || ||  |   |  |  ||      |||  |  | ||| ||||         |
Db        1703 GCGCACCACCTCCTCGGCGGCGTCAGCCCCGACGCCGACTTCACCGCGGAGGACTTCCGC 1762

Qy         303 GGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCC 362
                 |   | |||  |  || | ||  |  ||||  |  | |||  | | ||  ||| | ||||
Db        1763 CGCGAGGCGGCCGCCGCCGTCGCCCGCGTCCACGCGGCCGGCCGCCTCCCCGTCGTCGCC 1822

Qy         363 GAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTG 422
               |    |    | ||| |  |  |  |    |||||  |||||  |  ||  |||  || |
Db        1823 GGCGGGTCGAACATCTACGTCGAGGCGCTCGTGGCCGGCGGCGGCGGCGCGTTCCTCGCG 1882

Qy         423 TCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAG 482
               | | ||  |  |      |  | || |  |  |||||  ||  | | ||||     ||  |
Db        1883 GCGTACGACTGCCTCTTCCTGTGGACCGACGTCGCGCCGGACCTGCTGCGGTGGTACACG 1942

Qy         483 GGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGC 542
               | ||  | ||  || ||   || || || |   | | | |  ||   ||         ||
Db        1943 GCGGCGCGCGTGGACGACATGGTGCGGCGCGGGCTGGTTGGCGAGGCCC---------GC 1993

Qy         543 TCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCAC 602
                 || ||  || ||| |  |  | | |||| || | |  |     ||| | |  |  ||
Db        1994 GCCGGGTTCGACGCCGGGGCGGACTACACCCGCGGCGTGCGCCGCGCCATCGGGCTACCC 2053

Qy         603 GGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAG 662
               |         |  |||   ||||      | ||||  ||||  |    ||  ||
Db        2054 GAGATGCACGGCTACCTGCTGGCGGAGCGCGAGGGCGGCGCCGGCGCCGAGGACGACGAC 2113

Qy         663 ATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTG 722
                 | |||| | |   | | | | |  | | |   |   |       | | |  ||
Db        2114 GACCTCCTCGCCGGCATGCTCGAGGCCGCCGTGCGCGAGATCAAGGACAACACGTTCCGC 2173

Qy         723 CTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCC 782
               |||  ||  | | |    || | || |||  ||   ||    | | |  | | |  |
Db        2174 CTCACCGTGTCGCAGGTGGCCAAGATCCGGCGCCTCAGCGCGCTGCCCGGGTGGGACGTC 2233

Qy         783 TATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTC 842
                   |       |||    |   ||| |||  |      |    || | |    |
Db        2234 CGGCGCGTGGACGCGACGGCGGTGGTGGCGCGCATGGCGGAGGGCGCGCCCCACGGCGAG 2293

Qy         843 GCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACG 902
                 | |       |  |  ||  | ||||   |    ||  | |||   ||  | ||  |||
Db        2294 ACGTGGAGGGAGGTGGTGTGGGAGCCGTGCGAGGAGATGGTCAGCCGCTTCCTCGAGACG 2353

Qy         903 CTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCG 962
                 |  ||||   |||||| |   ||  |  ||  |     ||  | | |||| ||  ||||
Db        2354 CCCGCCGCCGCCGCCGCCGTCGTTGCCAACGGCAAAGTCGACGTGAACGTCGGCGACGCG 2413

Qy         963 ACGGCGGACGTCGCGCCCACCCCCGCC 989
                 | || | | |     |  |||||
Db        2414 GCCGCCGGCTTGCCTGAGGCTGCCGCC 2440


RESULT 7
US-11-075-185-35
; Sequence 35, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
;   LENGTH: 10968
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-35

  Query Match             7.6%;  Score 75.8;  DB 14;  Length 10968;
  Best Local Similarity   47.3%;  Pred. No. 3.7e-07;
  Matches  372;  Conservative    0;  Mismatches  397;  Indels   18;  Gaps    4;

Qy           1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
               | ||||||   |  |||  ||  | |  |   |||  ||||||  |||||||| |||
Db        8018 TCGACTGGCTCCGCCTCTTCGACCCCTCCGGCGCGCGCCGCGTCGAGCTGCCCACCTACG 8077

Qy          61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
               |||||   ||  |||| ||||||  ||      |     |     |  || |||||     |
Db        8078 CCTTCCAGCGACAGCGCTACTGGCTCCAGGCGCCGAGGCCTCGCCCCAGCCTCGAGGGCG 8137
```

```
Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
Db       8138 TTGGCCTCACCGCCGCAAACCACCCATGGCTCGG---CGCAGCCGTCCGCCTCGCCGACC 8194

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
Db       8195 GCGATGGCTACGTCCTCAGCGGCCGCCTCTCCACACTCGACCACCCGTGGGTCCTGGACC 8254

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGT 300
Db       8255 ACGTGGTGGCAGGCACGGTGATCTTGCCAGGAACGGCGTTCGTCGACCTGGCGTGGGCGG 8314

Qy        301 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCG---CGCCCGTCGTCGTGG 357
Db       8315 CGGCCGAGGTGGTGGGCGCCGCCGCTGTGTCCGAGGTGACCTTCACGACGCCGCTCGTGC 8374

Qy        358 GCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCG---CCGGCGCCATCGACT 414
Db       8375 TTCCGCCGCGCAGCGTGGTGGAGCTGCAGGTGAGGATCGGCGAGCCGGACGCGTCCGGGC 8434

Qy        415 TCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCC 474
Db       8435 GGCGGACGTTCGCCGCGTACAGCCGCCCGGACGCGGCGAGCGAGGCGGAGTGGACGCAAC 8494

Qy        475 TCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCT 534
Db       8495 ACGCGACCGGCGTGCTGAGCGCGCAGGCGGCGGCCGGGGCCGACGTGGCGGACCTTTCGG 8554

Qy        535 CACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGC 594
Db       8555 TGTGGCCGCCGCCGGGCGCCGAGGTGGTGGCGCTCGACGGCGGCTACGCGTGGCTGGCGG 8614

Qy        595 TGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCA 654
Db       8615 CGCAGGGCTACGGCTACGGCCCGGCGTTCCAGGCGCTGCGCGAGGTGTGGCGCGCGGGCA 8674

Qy        655 TCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCAC---------CTTTG 705
Db       8675 CGACGCTGTACGCGCGGGTCGCGCTGCCGGACGCGGTGGCGGACACGGCGCAGAGCTTCG 8734

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
Db       8735 GGATCCATCCGGCGCTGCTCGACGCGGTGCTGCACTCGTTGCTGGCGCGGTCGCCGCAGG 8794

Qy        766 AGGCGGC 772
Db       8795 AGGAGGC 8801
```

RESULT 8
US-11-075-185-1
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 78869
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-1

Query Match 7.6%; Score 75.8; DB 14; Length 78869;
Best Local Similarity 47.3%; Pred. No. 2.6e-07;
Matches 372; Conservative 0; Mismatches 397; Indels 18; Gaps 4;

```
Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
Db      23285 TCGACTGGCTCCGCCTCTTCGACCCCTCCGGCGCGCCGCGTCGAGCTGCCCACCTACG 23344

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG 120
Db      23345 CCTTCCAGCGACAGCGCTACTGGCTCCAGGCGCCGAGGCCTCGCCCCAGCCTCGAGGGCG 23404

Qy        121 CGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACA 180
Db      23405 TTGGCCTCACCGCCGCAAACCACCCATGGCTCGG---CGCAGCCGTCCGCCTCGCCGACC 23461

Qy        181 CGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCC 240
Db      23462 GCGATGGCTACGTCCTCAGCGGCCGCCTCTCCACACTCGACCACCCGTGGGTCCTGGACC 23521

Qy        241 AGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGT 300
Db      23522 ACGTGGTGGCAGGCACGGTGATCTTGCCAGGAACGGCGTTCGTCGACCTGGCGTGGGCGG 23581

Qy        301 CCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCG---CGCCCGTCGTCGTGG 357
Db      23582 CGGCCGAGGTGGTGGGCGCCGCCGCTGTGTCCGAGGTGACCTTCACGACGCCGCTCGTGC 23641

Qy        358 GCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCG---CCGGCGCCATCGACT 414
Db      23642 TTCCGCCGCGCAGCGTGGTGGAGCTGCAGGTGAGGATCGGCGAGCCGGACGCGTCCGGGC 23701

Qy        415 TCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCC 474
Db      23702 GGCGGACGTTCGCCGCGTACAGCCGCCCGGACGCGGCGAGCGAGGCGGAGTGGACGCAAC 23761

Qy        475 TCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCT 534
Db      23762 ACGCGACCGGCGTGCTGAGCGCGCAGGCGGCGGCCGGGGCCGACGTGGCGGACCTTTCGG 23821

Qy        535 CACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGC 594
Db      23822 TGTGGCCGCCGCCGGGCGCCGAGGTGGTGGCGCTCGACGGCGGCTACGCGTGGCTGGCGG 23881

Qy        595 TGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCA 654
Db      23882 CGCAGGGCTACGGCTACGGCCCGGCGTTCCAGGCGCTGCGCGAGGTGTGGCGCGCGGGCA 23941

Qy        655 TCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCAC---------CTTTG 705
Db      23942 CGACGCTGTACGCGCGGGTCGCGCTGCCGGACGCGGTGGCGGACACGGCGCAGAGCTTCG 24001

Qy        706 CCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATG 765
Db      24002 GGATCCATCCGGCGCTGCTCGACGCGGTGCTGCACTCGTTGCTGGCGCGGTCGCCGCAGG 24061

Qy        766 AGGCGGC 772
Db      24062 AGGAGGC 24068
```

RESULT 9
US-10-496-351-1/c
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712

```
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             7.4%;  Score 74.2;  DB 7;  Length 14055;
  Best Local Similarity   44.9%;  Pred. No. 7.5e-07;
  Matches  448;  Conservative    0;  Mismatches  538;  Indels   11;  Gaps    4;

Qy          6 TGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTC 65
              |||| |||| | |  ||||| | || |    |   ||  |  |||  |  | ||  |
Db      11822 TGGGCGCCGGTGGCGGCGCGGCCGGTGAACGCGGCGGTGTCCGAGCCGGGGATCTCCAGC 11763

Qy         66 GCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG---CG 122
               |||  | | |   | ||| ||||||  | |  |   |  |       |||   ||   ||
Db      11762 TCCCCGGTGAGCAGCGGGTGCTCCAGCGGGGCGGCGTCGGTGGTGGGCGGGGCGGGGTCG 11703

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |  ||| |||   || |  |  || | ||| |  ||  ||   | |||   ||| | ||
Db      11702 AGCCAGAAGCGGCGGCGCTGGAACGGGTAGGTGGGCAGGTCGGTGGTCCGGCCGCCCGCG 11643

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              |    | ||| ||  |||||||| |    |   | ||||  || | |   || || | ||
Db      11642 T----CGGCCAGCAGCGGGGCGAAGTCGACGCGCGCGCCCCTCGCGTGCAGCGCGCCGAG 11587

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
                  |  | | ||  || |  | | ||          ||||     | | |
Db      11586 CGCGGTGATCAGGGAGTGGTCCTCGTCCCGGTCGCGGTGCAGCAGGGGGGCGCAGAGCAC 11527

Qy        303 GGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCCGTCGTCGTGGGCGCC 362
              |    ||| ||||  |   ||||   || || | |      |  ||  ||  |  | |  |
Db      11526 GTCGTGCCCGTCCAGGCACTCCTGCGCGGCGGCGGCCAGCCCACCACCGGGGCCGACCTC 11467

Qy        363 GAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTG 422
              ||    ||||    ||       || | | |    |  | |||||  |  |||      | |
Db      11466 GACGAACGCGGTCGCGCCCCGCCGGTGCAGGTCGGTGACGGCGTCGGCGAAGCGCACCGG 11407

Qy        423 TCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG-- 480
              | | | |   |  | ||  ||   | ||  ||  ||  |       ||||    | |
Db      11406 CTCGCGGGCCTGGCCCAGCCAGTACTCGGGCGCGTCGAGCGCGCCGGGCGCGGTGGTGGC 11347

Qy        481 -AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGT 539
               |  ||||  |      | ||  |  | | | | || || ||  | |    ||   |
Db      11346 CACCGGGACGCGCGGCGGGCGGAAGTCCAGGCGCGCCAGCACGTCGGCGAACGCGGGCAG 11287

Qy        540 TGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCAC 599
              |      | |  |||| |||  |   ||    ||  ||  |||  ||  ||   || |
Db      11286 CACCGCGTCGACGGCGGCCGAGTGGAAGGCGTGGGCGACCCGCAGGCGCGTGGCGCGCCC 11227

Qy        600 CACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCG 659
              | |  ||| ||  |  ||    |  ||  ||    |  || |||  ||  | |   | |
Db      11226 GAGGCGGGCGGCGACCTCGGTGACGGCGGCCTCGTCGCCGGTGAGCACGGTGGCGCGGGG 11167

Qy        660 GAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCC 719
              |   ||  |  |    |||| |||| ||| |||      |  |    |||    | ||||
Db      11166 GCCGTTGACGGCGGCCAGCTCCACCCCGTCGACCAGCGCGGCGCGCGCCTCGTCCTCGCC 11107

Qy        720 CTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCC 779
              |     |||  | ||  ||   |  |      |||  | | || |  || || || | |
Db      11106 CGCGCGGACCGCGACCATCACGCCGGGGGGCAGCTCGCGCATGAGGGCGCCGCGCGCGGC 11047

Qy        780 CCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTC 839
              | | |   |  || | ||  |      |  | || ||   |    |  |||
Db      11046 CACGAGGGTGGAGGCGTCCTCCAGGGACAGCACCCCGGCGGCGTGCGCGACGGCGATCTC 10987

Qy        840 TTCG-CCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGT 898
                || || |  | || ||       || || |  ||| |||  |  |  | |
Db      10986 GCCGACCGAGTGCCCGGCGAGCAGGTCGGGGCGGACGCCGAGCGCGTCGAGCAGGCGCAC 10927

Qy        899 CACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCT 958
              ||    | ||    | ||   | ||  |  |  || |  ||| |  |   |   | |||
Db      10926 CAGCGCGACCTGCACGGCGAACAGGCCGGCCTGCGCGTAGCGGGTGTCGGCGAGCAGCCC 10867

Qy        959 CGCGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGG 995
               | | || |   ||  |||| ||| || |  | || |
Db      10866 GGGGTCGTCCAGCGCGGCGCGCACGCCGGGACCGCCG 10830


RESULT 10
US-10-496-351-56
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
;   LENGTH: 82746
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

  Query Match             7.4%;  Score 74.2;  DB 7;  Length 82746;
  Best Local Similarity   44.9%;  Pred. No. 5.5e-07;
  Matches  448;  Conservative    0;  Mismatches  538;  Indels   11;  Gaps    4;

Qy          6 TGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTC 65
              |||| |||| | |  ||||| | || |    |   ||  |  |||  |  | ||  |
Db      25220 TGGGCGCCGGTGGCGGCGCGGCCGGTGAACGCGGCGGTGTCCGAGCCGGGGATCTCCAGC 25279

Qy         66 GCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGG---CG 122
               |||  | | |   | ||| ||||||  | |  |   |  |       |||   ||   ||
Db      25280 TCCCCGGTGAGCAGCGGGTGCTCCAGCGGGGCGGCGTCGGTGGTGGGCGGGGCGGGGTCG 25339

Qy        123 GGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACG 182
              |  ||| |||   || |  |  || | ||| |  ||  ||   | |||   ||| | ||
Db      25340 AGCCAGAAGCGGCGGCGCTGGAACGGGTAGGTGGGCAGGTCGGTGGTCCGGCCGCCCGCG 25399

Qy        183 TACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAG 242
              |    | ||| ||  |||||||| |    |   | ||||  || | |   || || | ||
Db      25400 T----CGGCCAGCAGCGGGGCGAAGTCGACGCGCGCGCCCCTCGCGTGCAGCGCGCCGAG 25455

Qy        243 AAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCC 302
                  |  | | ||  || |  | | ||          ||||     | | |
Db      25456 CGCGGTGATCAGGGAGTGGTCCTCGTCCCGGTCGCGGTGCAGCAGGGGGGCGCAGAGCAC 25515

Qy        303 GGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCCGTCGTCGTGGGCGCC 362
              |    ||| ||||  |   ||||   || || | |      |  ||  ||  |  | |  |
Db      25516 GTCGTGCCCGTCCAGGCACTCCTGCGCGGCGGCGGCCAGCCCACCACCGGGGCCGACCTC 25575

Qy        363 GAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTG 422
              ||    ||||    ||       || | | |    |  | |||||  |  |||      | |
Db      25576 GACGAACGCGGTCGCGCCCCGCCGGTGCAGGTCGGTGACGGCGTCGGCGAAGCGCACCGG 25635

Qy        423 TCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACG-- 480
              | | | |   |  | ||  ||   | ||  ||  ||  |       ||||    | |
Db      25636 CTCGCGGGCCTGGCCCAGCCAGTACTCGGGCGCGTCGAGCGCGCCGGGCGCGGTGGTGGC 25695

Qy        481 -AGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGT 539
               |  ||||  |      | ||  |  | | | | || || ||  | |    ||   |
Db      25696 CACCGGGACGCGCGGCGGGCGGAAGTCCAGGCGCGCCAGCACGTCGGCGAACGCGGGCAG 25755

Qy        540 TGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCAC 599
              |      | |  |||| |||  |   ||    ||  ||  |||  ||  ||   || |
```

```
Db      25756 CACCGCGTCGACGGCGGCCGAGTGGAAGGCGTGGGCGACCCGCAGGCGCGTGGCGCGCCC 25815

Qy        600 CACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCG 659

Db      25816 GAGGCGGGCGGCGACCTCGGTGACGGCGGCCTCGTCGCCGGTGAGCACGGTGGCGCGGGG 25875

Qy        660 GAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCC 719

Db      25876 GCCGTTGACGGCGGCCAGCTCCACCCCGTCGACCAGCGCGGCGCGCGCCTCGTCCTCGCC 25935

Qy        720 CTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCC 779

Db      25936 CGCGCGGACCGCGACCATCACGCCGGGGGGCAGCTCGCGCATGAGGGCGCCGCGCGCGGC 25995

Qy        780 CCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTC 839

Db      25996 CACGAGGGTGGAGGCGTCCTCCAGGGACAGCACCCCGGCGGCGTGCGCGACGGCGATCTC 26055

Qy        840 TTCG-CCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGT 898

Db      26056 GCCGACCGAGTGCCCGGCGAGCAGGTCGGGGCGGACGCCGAGCGCGTCGAGCAGGCGCAC 26115

Qy        899 CACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCT 958

Db      26116 CAGCGCGACCTGCACGGCGAACAGGCCGGCCTGCGCGTAGCGGGTGTCGGCGAGCAGCCC 26175

Qy        959 CGCGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGG 995

Db      26176 GGGGTCGTCCAGCGCGGCGCGCACGCCGGGACCGCCG 26212


RESULT 11
US-10-496-351-22
; Sequence 22, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 2742
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-22

  Query Match             7.2%;  Score 72.2;  DB 7;  Length 2742;
  Best Local Similarity   44.7%;  Pred. No. 2.6e-06;
  Matches  368;  Conservative    0;  Mismatches  448;  Indels    7;  Gaps    2;

Qy         12 CCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCGC 71

Db        764 CGGCCCGGCGCGAGGCGGCGCCCGCCGCGCGGTCCGGGGCGACCGAGGGCGGTGGCCCGG 823

Qy         72 GAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAGG 131

Db        824 TCGCGGCGCCCCCGGTCGCGGCGC----CCCCGGCCGGGATCGTCCCGCAGCGGCAGGCG 879

Qy        132 CGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGC 191

Db        880 CCCCCGGCGCCCGCGCCTGCCCCCGTCGCGCTCGCGCCCCTCCCCCCGGCGTCCGCGCCG 939

Qy        192 CGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTC 251

Db        940 ACCGCCGAGCTCAAGCACGTCAGCGTCCTCATGGTCCTCACCCGGTTCGAGCACGACGTG 999

Qy        252 CCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCG 311

Db       1000 GCCGCCGCCCCGCCCGAGGAGGTGGACGACGCGCTGAGCGACGTCGCCACCGCGGTGCGC 1059

Qy        312 GTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGC 371

Db       1060 GAGGAGGTCGAGCGGGCGGGCGGCGTCGTCACCGCCAAGGCCGCGTCGCTGATCGTCGCC 1119

Qy        372 GAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCC 431

Db       1120 CTGTTCGGCTCCACCCGCAGCGGCGAGGACGACGCCACCAGGGCGGTGCGCGCCGCGCTC 1179

Qy        432 GCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCC 491

Db       1180 GCCGTGCGCGACCGGCTCGCCCACGGGCCGCGCGGCGCCGGGCCCGCCGCGCTGGGCGTC 1239

Qy        492 GCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTC 551

Db       1240 ACCGCCGCCGTCGCCACCGGGCACGCCCTGGTGCGGCGCCACCCAGGGCGCGCGCCCGAG 1299

Qy        552 ACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGC 611

Db       1300 GTCGGCGGCGGCGTGCTCGACCGCTGCATGAGGCTGCTCGCGCACGTCTCGCCCGGCGAG 1359

Qy        612 ATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTC 671

Db       1360 GTCAGGGTGTGCGAGGAGACCAGGGGCGCCAGCGAGGGCGCCATCGGCTACGACGGCCCC 1419

Qy        672 CCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCT 731

Db       1420 CACGACCCGGTGCGCGGGTCCTGGGCCACCGGGGCCAGGCAGGGCGCGGGCGGGCCCCCG 1479

Qy        732 GCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCT 791

Db       1480 GCCACGC---CGCCGCTGGTGGACCGGGACCGCGAGCTGGCCCTGCTGCGCGGCCTGCTG 1536

Qy        792 TTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGC 834

Db       1537 GAGCAGGTCGGCCGCAGGTCCCGGACGCACCTGGTGACCGTGC 1579


RESULT 12
US-11-096-568A-14645
; Sequence 14645, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides I
;  TITLE OF INVENTION:  Therby
;  FILE REFERENCE: 2750-1592PUS2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE:  2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 14645
;   LENGTH: 2489
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2489)
;   OTHER INFORMATION: Ceres Seq. ID no. 11049475
US-11-096-568A-14645

  Query Match             7.1%;  Score 71.2;  DB 11;  Length 2489;
  Best Local Similarity   46.5%;  Pred. No. 4.3e-06;
  Matches  334;  Conservative    0;  Mismatches  378;  Indels    6;  Gaps    3;

Qy         20 CGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCGCGAGCGTTA 79

Db        210 CGTGCCGCCCGCCGAGTGCCTCCAGGCGCCGTTCCACAAGCCCGTCGCCGACCTGCCTGC 269

Qy         80 CTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAGGCGCTTGCA 139

Db        270 GCCGTCCAGGAAGCCCGCCGCCATTGCCGTCCCAGGGCACGCCGCGGCGCCGAGGAAAGC 329
```

```
Qy        140 CCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGCCGGCTCCG 199
              |  ||   ||  | |   | |  ||||||      | ||    ||   |    |
Db        330 GGAGGGCGGCAAGAAGCAGCTCAACTTGTTCCAGCGCGCCGCGGCGGCCGCGCTCGACGC 389

Qy        200 GGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGC 259
              |  ||| |     | | | ||| |   ||||| ||    |||   ||| | |  | ||||
Db        390 GTTCGAGGAAGGGTTCGTGGCCAACGTCCTCGAGCGGCCCCACGGGCTGCCCAGCACGGC 449

Qy        260 GG---CGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCG 316
               |   ||  | || |||| ||| ||  |  || |    | | |    ||| ||| || |
Db        450 CGACCCGGCCGTGCAGATCGCCGGCAACTTCGCGCCCGTCGGGGAGAGGCCGCCCGTGCA 509

Qy        317 CGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGAT 376
              |||  ||| || |  || | | | | ||||| || |||  || | | |  ||||| |
Db        510 CGAGCTCCCCGTCTCCGGCCGCATCCCGCCCTTCATCGACGGGGTCTACGCGCGCAACGG 569

Qy        377 CGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGA 436
              ||  |   |   |       ||| ||  |    | || ||  ||    |  || |   |
Db        570 CGCCAACCCCTGCTTCGACCCCGTCGCGGGGCACCACCTCTTCGACGGCGACGGCATGGT 629

Qy        437 GCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGA 496
              || |||| ||       | |   | | | ||       |||   |  |  |  || ||  ||
Db        630 GCACGCGCTGCGGATACGC-AACGGCGCCGCCGAGTCCTACGCCTGCCGCTTCACGGAGA 688

Qy        497 AGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGG 556
              |||  |  ||||     ||| ||||  |  || | |    | | | | || ||  ||
Db        689 CCGCGCGCCTGCGCCAGGAGCGCGCGAT--CGGCCGCCCCGTCTTCCCCAAGGCCATTGG 746

Qy        557 CGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCG 616
              ||          |  ||  | |  ||  ||  ||||  |||  | |||   ||  |  | |
Db        747 CGAGCTGCACGGGCACTCCGGGATCGCGCGCCTCGCCCTGTTCTACGCGCGCGCCGCGTG 806

Qy        617 CGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGA 676
              ||  |||   || |          |||   |||| ||||| ||| |   |   | |   |
Db        807 CGGCCTCGTGGACCCCTCGGCCGGCACCGGCGTGGCCAACGCCGGCCTCGTCTACTTCAA 866

Qy        677 GCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCC 734
              | ||   | | | | |||     | |||||   |  | | | ||  || | |
Db        867 CGGCCGCCTGCTCGCCATGTCCGAGGACGACCTCCCCTACCACGTCCGCGTGGCGGAC 924


RESULT 13
US-11-205-109-1
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Staffa, Alfredo
;  TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
;  FILE REFERENCE: 3002-2US
;  CURRENT APPLICATION NUMBER: US/11/205,109
;  CURRENT FILING DATE:  2005-08-17
;  PRIOR APPLICATION NUMBER: US/09/976,059
;  PRIOR FILING DATE: 2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 88421
;   TYPE: DNA
;   ORGANISM: Actinoplanes sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2077)..(3078)
;   OTHER INFORMATION: ORF 1; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3118)..(4032)
;   OTHER INFORMATION: ORF 2; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4038)..(5048)
;   OTHER INFORMATION: ORF 3; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6665)..(5814)
;   OTHER INFORMATION: ORF 4; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7703)..(6693)
;   OTHER INFORMATION: ORF 5; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9464)..(8130)
;   OTHER INFORMATION: ORF 6; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9691)..(10761)
;   OTHER INFORMATION: ORF 7; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (12751)..(10829)
;   OTHER INFORMATION: ORF 8; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13617)..(12802)
;   OTHER INFORMATION: ORF 9; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15203)..(13614)
;   OTHER INFORMATION: ORF 10; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15591)..(15863)
;   OTHER INFORMATION: ORF 11; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15880)..(19035)
;   OTHER INFORMATION: ORF 12; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19032)..(39713)
;   OTHER INFORMATION: ORF 13; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (39713)..(65800)
;   OTHER INFORMATION: ORF 14; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65826)..(66530)
;   OTHER INFORMATION: ORF 15; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (66546)..(67370)
;   OTHER INFORMATION: ORF 16; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (67384)..(70059)
;   OTHER INFORMATION: ORF 17; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70099)..(70662)
;   OTHER INFORMATION: ORF 18; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70659)..(71906)
;   OTHER INFORMATION: ORF 19; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (73439)..(71964)
;   OTHER INFORMATION: ORF 20; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (74216)..(73563)
;   OTHER INFORMATION: ORF 21; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75424)..(74213)
;   OTHER INFORMATION: ORF 22; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75535)..(76464)
;   OTHER INFORMATION: ORF 23; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78110)..(76449)
;   OTHER INFORMATION: ORF 24; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79864)..(78107)
;   OTHER INFORMATION: ORF 25; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81624)..(79861)
;   OTHER INFORMATION: ORF 26; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81909)..(81682)
;   OTHER INFORMATION: ORF 27; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82346)..(82062)
;   OTHER INFORMATION: ORF 28; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82587)..(84446)
;   OTHER INFORMATION: ORF 29; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84481)..(85548)
;   OTHER INFORMATION: ORF 30; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85556)..(86845)
;   OTHER INFORMATION: ORF 31; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87372)..(86803)
;   OTHER INFORMATION: ORF 32; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87494)..(88420)
;   OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

  Query Match             7.1%;  Score 71.2;  DB 14;  Length 88421;
  Best Local Similarity   43.6%;  Pred. No. 2.3e-06;
  Matches  319;  Conservative    0;  Mismatches  413;  Indels    0;  Gaps    0;

Qy        208 CGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCC 267
              || || | | ||| | | | || | || | | | ||| ||| | ||||
Db      38179 CGATCGATCCCGCGTACCCGCGCGACCGCGTCGACTACCTCGTCCGGGACGCGGAGCCGG 38238

Qy        268 TGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCG 327
              |    ||     || ||||||| |    ||| | | || | |||  |  |   | |||
Db      38239 CGTGCCTGCTGACCGTCGCCGGGCATCGCGCGGCGGCGCCGGCGGCACCGGCGGTGGTCG 38298

Qy        328 ACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGC 387
              | | | |     | ||| |||  | || |  ||||||  |      |      |
Db      38299 AGCTGGACGACCCGGCGACCGCGGCCGAGATCGCCGATGCCGAACCGGAGCCGCCGGTCG 38358

Qy        388 TCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGA 447
              || | || ||  | ||  | || |   |    ||| | | |      |   |   |
Db      38359 CGGTGCGGCCCACCCACTCCGCCTACCTGATCTACACCTCGGGCTCGACCGGGCGGCCCA 38418

Qy        448 TCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGC 507
              |  || |      |   | | | | || |   || || | | | ||
Db      38419 AGGGCGTCGTGGTCACCCACCGGGGTGTGGCCGCGCTCGTGGCCACCCAGGCCGAGCGGC 38478

Qy        508 CGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGT 567
              | | |  |||   | | ||   | | | | || ||   |  || | | ||
Db      38479 TCGCGGTGACCGGCGAGAGCCGCGTCCTGCAGTTCGCCAGCGTGGGCTTCGACGCCTCGA 38538

Qy        568 GCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACG 627
              ||          |   |  ||  |     | | |     | |   | ||  | |||
Db      38539 TCTGGGAGATGGTGATGGCCCTGTGCGCCGGCGCCACCCTGGTCGTCGCCCCGGCCGACG 38598

Qy        628 AGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACG 687
              | ||||  |  |||  || |  ||||  | ||  |     | | | ||| |   |||
Db      38599 ACCTGCTGCCCGGCCCGGCCCTGGCCGCCACGCTGTCCGGGCACGCGGTGACCCACGCGA 38658

Qy        688 TGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGC 747
                  |  || |   |   |  |||| |    |||    || | ||    | ||
Db      38659 CCCTGCCGCCGGCCGTGCTCGCCGCGTCCGCGCCCGGCGATCTCGCGCCGCTCGCCGTGC 38718

Qy        748 TCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGG 807
              | |    |   |  || | |   |  | | | || |   |  |  |||||  ||||||
Db      38719 TGGTCTCGGCCGGTGAGGCGCTCGGGCCGGACCTCGTCCGGCAGTTCGCGCCCGGCCGCG 38778

Qy        808 TCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGC 867
              |  ||  |         |||| |   |  |   |  |  |||||| ||||  || |||
Db      38779 CGCTGGTGAACGCGTACGGGCCGACCGAGACCACGGTGTGCGCCACGGCGTCCGCCCCGC 38838

Qy        868 CGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGC 927
                    | |  ||  |  |     || | || | |||||| ||   | |   ||||
Db      38839 TCGGCCCGGAGGATCCCCCGCACATCGGCGCCCCGGTCGCCGATTCCCGGGTCTACGTGC 38898

Qy        928 TCCTCGAGATGC 939
              ||  |||   ||
Db      38899 TCGACGACGCGC 38910


RESULT 14
US-11-096-568A-27171
; Sequence 27171, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides F
;  TITLE OF INVENTION:  Therby
;  FILE REFERENCE: 2750-1592PUS2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE:  2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 27171
;   LENGTH: 1192
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1192)
;   OTHER INFORMATION: Ceres Seq. ID no. 15178969
US-11-096-568A-27171

  Query Match             7.1%;  Score 71;  DB 11;  Length 1192;
  Best Local Similarity   44.0%;  Pred. No. 5.4e-06;
  Matches  299;  Conservative    0;  Mismatches  380;  Indels    0;  Gaps    0;

Qy        242 GAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTC 301
              ||   ||  |  |  || || | |   | |||||  ||| | ||  |     ||||
Db      133 GACTTTCTCCGTCAAGGTGGGGCTGGCGCAGATGCTCCGGGGCGGGGTCATCATGGACGT 192

Qy        302 CGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGC 361
              | |  ||||  |  |    |||||  |  |      ||||  || |||  |||
Db      193 GGTGACCCCGGAGCAGGCGCGCCTCGCGGAGGAGGCCGGCGCGTGCGCCGTGATGGCGCT 252

Qy        362 CGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGT 421
              |||| | |     ||| ||       | |   | ||  |    |||     | | |
Db      253 GGAGCGGGTCCCCGCCGACATCCGCGCGCAGGGCGGCGTGGCGCGCATGTCGGACCCGGG 312
```

```
Qy        422 GTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGA 481
              |  |  || |          |   || ||  | ||     |   | |  || |  ||
Db        313 CCTCATCCGCGACATCAAGCGCGCCGTCACCATCCCGGTCATGGCCAAGGCCCGCATCGG 372

Qy        482 GGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTG 541
              |      |  |  ||    |  | || |     ||||  || | |   |  |
Db        373 GCACTTCGTGGAGGCCCAGATCCTGGAGGCCGTGGGCGTCGACTACGTGGACGAGAGCGA 432

Qy        542 CTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCA 601
              |  |  | ||| ||  |||||||  | ||  |  |   ||  |  || |  | | |
Db        433 GGTGCTGACCCCCGCCGACGACGCGCACCACATCAACAAGCACAACTTCCGCGTCCCCTT 492

Qy        602 CGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGA 661
              || | || ||   ||||  |||  ||||  || |||  | | | ||||||  |  |||
Db        493 CGTCTGCGGCTGCCGCGACCTCGGCGAGGCGCTCCGCAGGGTGCGCGAGGGCGCCGCCAT 552

Qy        662 GATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCT 721
              ||| |||  |  || |   |  |  || |  |  |  |  ||| | |  | ||  |
Db        553 GATCCGCACCAAGGGGGAGGCCGGCACGGGCAACGTCGTCGAGGCCGTGCGCCACGTGCG 612

Qy        722 GCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCC 781
              || |   | ||        || ||||| |     ||||||| || | || |   |||
Db        613 CTCCGTCATGGGCGACGTCCGCGCGCTCCGCAGCATGGACGACGACGAGGTCTTCGCCTA 672

Qy        782 CTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTT 841
              |      |  ||||||    |  | |   | | |   || |  | | | | |
Db        673 CGCCAAGCGCATCGCCGCGCCCTACGACCTCGTCATGCAGACCAAGCAGCTGGGCCGCCT 732

Qy        842 CGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCAC 901
              | ||   |   |   |  | | |      | ||||  ||||   ||| | | | | |
Db        733 CCCCGTCGTGCAGTTCGCCGCGGGGGGTGTCGCCACGCCCGCCGACGCCGCGCTCATGAT 792

Qy        902 GCTGGCCGATCCCGCCGGC 920
              || |  ||    || ||||
Db        793 GCAGCTCGGATGCGACGGC 811


RESULT 15
US-11-075-185-36
; Sequence 36, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36
;   LENGTH: 5679
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-36

  Query Match             7.0%;  Score 70.4;  DB 14;  Length 5679;
  Best Local Similarity   47.2%;  Pred. No. 5.4e-06;
  Matches  407;  Conservative    0;  Mismatches  421;  Indels   34;  Gaps    5;

Qy          1 TGGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACC 60
              | ||||||   |  |||  ||    |         ||| ||||||  ||||||||| |||
Db       2708 TCGACTGGCTCCGCCTCTTCGACGCCACCCGCGCGCGCCGCGTCGAGCTGCCCACCTACG 2767

Qy         61 CCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGC---------CTTCACCC 111
              ||||   ||| |||| ||||||  ||      |    |  |||          |  || |
Db       2768 CCTTTCAACGCCAGCGCTACTGGATCGATGCGCCGCGGCGCGCGGCGGGGCTCGACAGCG 2827

Qy        112 TCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCC 171
              |||| |  |||  |     | |     | |  || |  || |     | ||
Db       2828 TCGGGCTCACGGCCGCAGATCACCCCTGGCTGGGCGCGGCGGTGCGGCTCGCCGACCGGG 2887

Qy        172 ACGCGCACACGTACTCCAGCCGGCT---CCGGGGCGACGCGTTCTACCTCGCCGATCACC 228
              |||  |||  |     | | |||||       || ||||      |   ||  || |||
Db       2888 ACGTCCACGTGCTGAGCGGGCGGCTGTCGACGGTCGACCACCCGTGGATCCTGGACCACG 2947

Qy        229 TCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCG 288
              | |||   |||  |  |||  | ||  |   |||  ||| | ||| ||||  | ||  | |
Db       2948 TGGTGGCGGGCACGCCGCTGATGCCAGGAACGGGCTTCGTCGAGCTGGCGTGGGCGACGG 3007

Qy        289 GGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCG 348
              ||  || |  ||       || |||   ||  ||  |              |  || |
Db       3008 CCCAGGCGGTGGACGCCGCCGCGATCGCGGAGCTCACCCTGACGACGCCGCTCGTGTTGC 3067

Qy        349 TCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATA------------ACGCTCCGGCCGG 396
              |      |||||  | || || | | |  | |                ||    | | |
Db       3068 CGGCGCGCGGCGCGGTGCAGCTCCAGGTGACGGTCGACGAGGCCGACGCGAATGGCCGGC 3127

Qy        397 CCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACG 456
              ||   ||||||| ||  |      |   | ||| ||| | ||||  ||     ||||
Db       3128 GGGCATTCGCCATCCACAGCCGGCCGCATGGCCCCGGCGACCTCGCGTGGACGCAACACG 3187

Qy        457 CGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCC 516
              ||   ||   | || ||   || |||  ||      ||||| |||||||||
Db       3188 CGACCGGCGTGCTGAGCGCGGAGGAGCCGGCGGGAGCCGACGAGGCGGCG---------G 3238

Qy        517 CGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCC 576
              ||||   ||     |  | |   || ||  ||      |     ||||   | |||  |
Db       3239 GGCTCTCGGAGTGGCCGCCGCCGGGCGCGGAGGCGGTGGCGCTCGACGGCGGGTACGAGC 3298

Qy        577 GCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCC 636
              ||     ||| |||      |||||||       | |  | ||||| || |||  | |
Db       3299 AGCTGTCCGAGCACGGCTACGGCCACGGCCCGGCGTTCCAGGGGCTCCGCGGGCTCTGGC 3358

Qy        637 GAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCT 696
              | |  ||  |   |     |||| |  |    || |||||   |  |  |    |
Db       3359 GCGCGGACCGTACGCTGTACGCGCACGTCGCGCTGCCGGACGCTGTCGCGGGCACCGAGC 3418

Qy        697 CCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGC 756
              |||  |  |||||||| || ||  |||| || || |||||  |    |     |||||
Db       3419 AGGGCTTCGGGCTCCATCCGGCGCTCTTCGATGCGGCGCTGCAGTCGCTGGCGCGGCTGT 3478

Qy        757 TGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCG-GCCGGGTCACGCTG 815
              | ||| || ||||||  | |  |   | |  |  |  |||  | |   |||  |  |
Db       3479 CGCGCGAGGAAGCGGCCGCTGGCGACCCGGTGCTGGTGCCGTTCGCGTGGACGGACGTGG 3538

Qy        816 CTCCGCCCGCTGCCGGCGCGGC 837
              | | |  |||  |||||||| |
Db       3539 CGCTGTACGCGACCGGCGCGAC 3560


Search completed: April 27, 2006, 15:52:59
Job time : 628.947 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 4956.14 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_35000_35999
Perfect score: 1000
Sequence: 1 tggactgggagccgctcgac.....acccccgcccagcggctcta 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	113.2	11.3	1956	10	CG754548	CG754548 P050-1-A1
2	110	11.0	1798	10	AG171124	AG171124 Pan trogl
c 3	106.6	10.7	1628	10	CG757066	CG757066 P052-2-A0
4	106.4	10.6	1538	10	AG030607	AG030607 Pan trogl
c 5	105	10.5	1878	10	AG393083	AG393083 Mus muscu
c 6	102.8	10.3	1189	10	AG030608	AG030608 Pan trogl
c 7	100.6	10.1	1552	10	AG430101	AG430101 Mus muscu
c 8	99.8	10.0	1961	10	AG435522	AG435522 Mus muscu
c 9	99.4	9.9	1531	10	AG430426	AG430426 Mus muscu
c 10	99	9.9	1341	10	AG030611	AG030611 Pan trogl
11	98.2	9.8	1159	9	CC210735	CC210735 CH261-185
12	98.2	9.8	1674	10	CL078342	CL078342 CH216-149
c 13	97.6	9.8	2243	10	AG381986	AG381986 Mus muscu
c 14	97.2	9.7	1462	10	AG441877	AG441877 Mus muscu
c 15	97.2	9.7	1822	10	AG435170	AG435170 Mus muscu
c 16	94.4	9.4	1280	10	CL470317	CL470317 SAIL_140_
c 17	94.4	9.4	1360	10	CL469817	CL469817 SAIL_133_
c 18	94	9.4	1506	10	AG442117	AG442117 Mus muscu
c 19	93.8	9.4	1015	9	BZ569259	BZ569259 pacs2-164
c 20	93.8	9.4	1302	10	AG448240	AG448240 Mus muscu
c 21	93.6	9.4	1406	10	AG441598	AG441598 Mus muscu
c 22	93.2	9.3	1523	10	AG448267	AG448267 Mus muscu
c 23	93.2	9.3	1821	10	CL090560	CL090560 ISB1-17N1
c 24	92.8	9.3	1551	10	AG341259	AG341259 Mus muscu
25	92.6	9.3	1406	10	CG756569	CG756569 P051-4-B0
c 26	92.6	9.3	1440	10	AG332984	AG332984 Mus muscu
c 27	92.4	9.2	1450	10	AG429738	AG429738 Mus muscu
c 28	91.8	9.2	1355	10	AG429718	AG429718 Mus muscu
c 29	91.4	9.1	1147	1	AJ928063	AJ928063 AJ928063
30	91.4	9.1	1392	10	AG390642	AG390642 Mus muscu
c 31	90.8	9.1	1569	10	AG341503	AG341503 Mus muscu
c 32	90.6	9.1	1675	10	AG360980	AG360980 Mus muscu
c 33	90	9.0	1116	9	BZ569478	BZ569478 pacs2-164
c 34	90	9.0	1661	10	AG435089	AG435089 Mus muscu
c 35	90	9.0	1885	10	AG363563	AG363563 Mus muscu
c 36	89.4	8.9	1309	10	AG077201	AG077201 Pan trogl
c 37	89.4	8.9	1453	8	DN806685	DN806685 76982446
c 38	89.4	8.9	1798	10	AG171124	AG171124 Pan trogl
c 39	89.2	8.9	1137	2	BG809979	BG809979 mgct002xd
c 40	88.6	8.9	2017	10	AG435283	AG435283 Mus muscu
41	88.6	8.9	2420	10	CL509247	CL509247 SAIL_811_
c 42	88.4	8.8	1651	2	BG809816	BG809816 mgct001xk
43	88.4	8.8	1843	9	BZ557952	BZ557952 pacs1-60_
c 44	88.2	8.8	1037	10	AG061836	AG061836 Pan trogl
c 45	87.8	8.8	1425	10	AG441757	AG441757 Mus muscu

ALIGNMENTS

RESULT 1
CG754548/c
LOCUS CG754548 1956 bp DNA linear GSS 24-OCT-2003
DEFINITION P050-1-A11.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG754548
VERSION CG754548.1 GI:37980151
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1956)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1956
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 11.3%; Score 113.2; DB 10; Length 1956;
Best Local Similarity 44.0%; Pred. No. 6.1e-12;
Matches 440; Conservative 0; Mismatches 552; Indels 8; Gaps 2;

Qy 7 GGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCG 66

```
Db       1865 GCGGGCCCCCCGGCCCGCGGCGCGGCCGGCGCGGGCCGGCCGGCGCGGGCGNCCGCCGCG 1806

Qy         67 CCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAG 126

Db       1805 CGCCCCCGGGGGCNGGCGCCNGCCNGCGCGGGCGCCCCCGCCGCCCCGCCCCGCCGCGC 1746

Qy        127 AGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACT 186

Db       1745 GGCGCCCCCCGCCGCCGCGCCGGCGCCCCCGCCCCGGCCCGCCGCCGCCGCCCCCGCGGG 1686

Qy        187 CCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGC 246

Db       1685 CGGGCCCCCCCCGCCCCCCCCGCGCCGCCCGCCCCCGCCGCCCCGCCNCCCCCCGCGGCC 1626

Qy        247 TCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGG--GGAGATGGCGTCCGG 304

Db       1625 CCCGGGGCGGCCCCGCGGCGCCCGGGCCGCCCGGCCCCGCGGCCCGGCGGCGNCGCGGCC 1566

Qy        305 GCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGA 364

Db       1565 GCGGGGGGGGCGGCCCCCCCGCGGCGGCCCCCGGGGGGGGCGGGGGGCGCGGCGNNGCCC 1506

Qy        365 GCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTC 424

Db       1505 GGGCCGCGCCCCCCGGGCCGGGCCCCGGCGCGCCGCCGCGGCCCCCCCGCGCGCCGCCCC 1446

Qy        425 CTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGG 484

Db       1445 CCGCCCCCCCGCCCCCGCGCCCCNGCGCGCCGGCCCCCGGGCCCCCGCGCCGGCCCCGCC 1386

Qy        485 GGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTC 544

Db       1385 GCCCCCGCCCCGGCGCGCCGCCCCCCCCCCCCGGCGCCCCCCGCCCCCCGCCCCCGCGGC 1326

Qy        545 GAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGG 604

Db       1325 CCCCGCCCGCGGGCGCGCCCGGGNGCCCNGGNGCCCCCGCGGGCCGCCGCCCCGCGGNGG 1266

Qy        605 CGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGAT 664

Db       1265 CGGCCCCGNCCGGGCGGCCCCCGCGCCCCGNCGCGCCCCGCGCGCCCGCGGGGCCGNNGC 1206

Qy        665 TCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCT 724

Db       1205 GGCCCGCNCGGCCGCGCCCCGCGCGGGGCCCGNCCCCCGCGGCCGCGCGCCGCCCCGGCC 1146

Qy        725 CGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTA 784

Db       1145 GCGCGGCCCCCCCGCCCGCCCGGCNGCCGCCCGCCCGCCCCCGGCNGNGCGGGCGGCGCG 1086

Qy        785 TCTTCCTTTCGCCGTCGGCC------GGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCT 838

Db       1085 CCGGCCNGGCGGCCCCGGCCCGCGCGGGCCCCCGCCGCGGGCCGGGCCGCCCCCGCGNCG 1026

Qy        839 CTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGT 898

Db       1025 CGGCCCGCCGCGCCCGCCCGNGCGCCGGCCGCCGCGCGCGNGCGCCGGCNCCCCGCCCGG 966

Qy        899 CACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCT 958

Db        965 CGCGNCCCGCCCGCCCNNCCCNCNNCNGNCCCGCGGCGGGCCCCCNNCCCNCCNCCCCCC 906

Qy        959 CGCGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTC 998

Db        905 CCCCCCCNNNNNNCCCCCNNNCCNNCCCCCCCCNNNNCCC 866


RESULT 2
AG171124
LOCUS       AG171124               1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                Sequencing: TJ
             LIBRARY
                Vector    : pBACe3.6
                R.Site 1  : EcoRI
                R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match             11.0%;  Score 110;  DB 10;  Length 1798;
  Best Local Similarity   43.1%;  Pred. No. 2.5e-11;
  Matches  429;  Conservative    0;  Mismatches  561;  Indels    6;  Gaps    4;

Qy          7 GGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCG 66

Db        786 GGCGCCCGCNCNCCGCGCGCCGCCGCGGCGCNNNGCGCGCCGCGCNCGCGCCGCGCGNCG 845

Qy         67 CCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAG 126

Db        846 CCCCCGCGGCGCGCNCGCCCGCGNCGCCGCGCGCGCNCGCCGCGCCGCGCGCCNCGCGCC 905

Qy        127 AGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACT 186

Db        906 CGCGCGNNCCGCGGCGCCGCGCCCGCCNCGCCCGCNCGGCGCNCNNCGCCGCCCCGCGCC 965

Qy        187 CCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGC 246

Db        966 NCNCGCCGCCGCGGCGCNCGCCGCCCGCGCGCCGCCGCCCNCGGCGC-GCCNGCNGC 1024

Qy        247 TCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGC 306

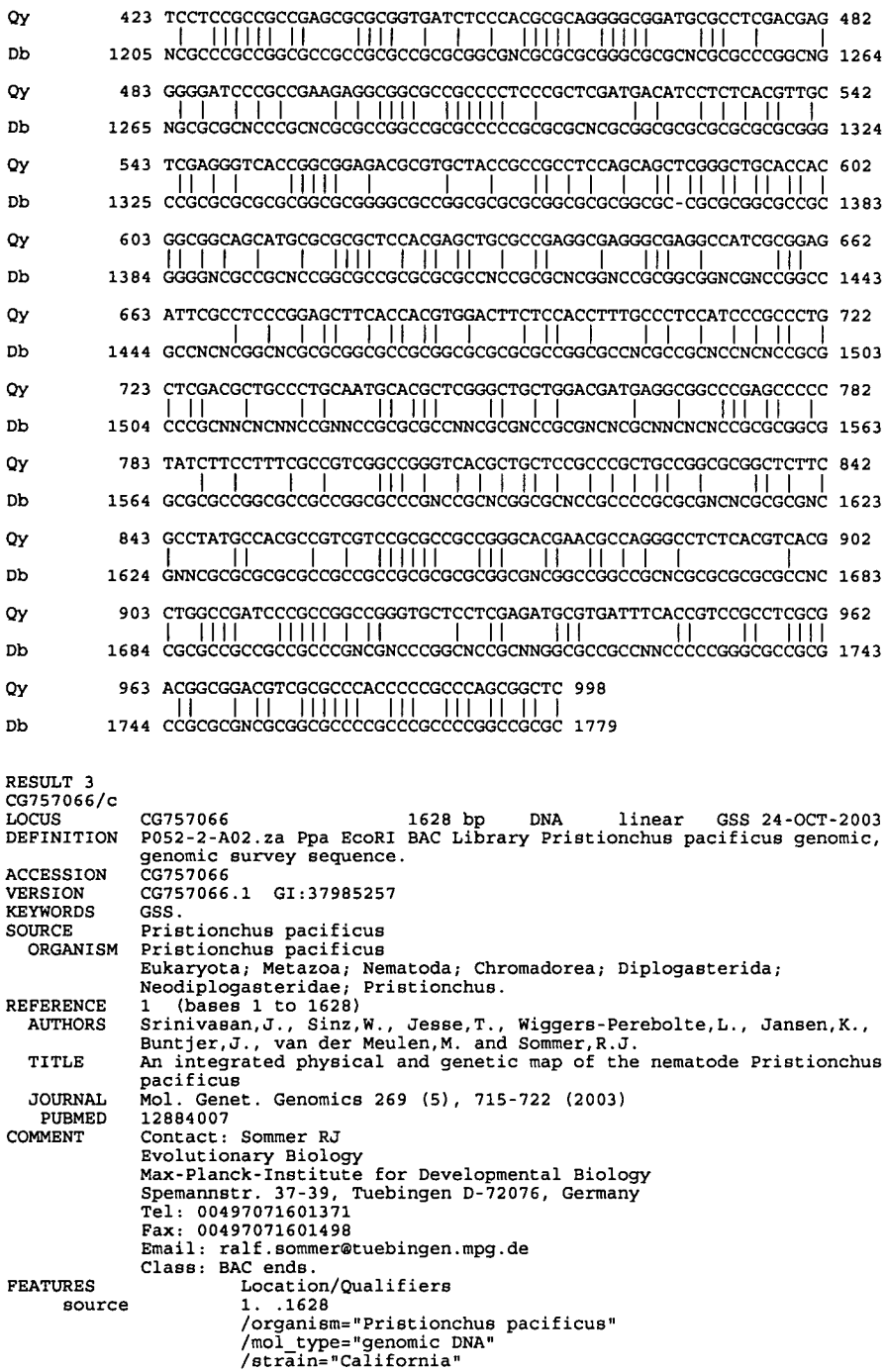
Db       1025 NCGCCCGCGCGCCNNNNNNGCGCCGGCGCGCGCGCGCGNCGCGCGCGCGCNNNCGCGCC 1084

Qy        307 GGCCGGTCCGCGACATCCTCGACGTCG-TCTGGACCGCGCCCGTCGTCGTGGGCGCCGAG 365

Db       1085 CGCGNCNCNCCGCGNCGCCCGCCCGCGCGCGCGNCCGCGGGCCGCCNCNCGCCCGCCGCG 1144

Qy        366 CCGCGCGAGATCGAGATAACGCTC---CGGCCGGCCGCCGGCGCCATCGACTTCGCCGTG 422

Db       1145 CNNCCGCNCNGCGGCGGCGCGCGCGCGCGCGCGNCGGCGCGCGCCGGCGCGCNCGCGGCN 1204
```



```
Qy        423 TCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAG 482
Db       1205 NCGCCCGCCGGCGCCGCCGCGCCGCGCGGCGNCGCGCGCGGGCGCGCNCGCGCCCGGCNG 1264

Qy        483 GGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGC 542
Db       1265 NGCGCGCNCCCGCNCGCGCCGGCCGCGCCCCCGCGCGCNCGCGGCGCGCGCGCGCGGG 1324

Qy        543 TCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCAC 602
Db       1325 CCGCGCGCGCGCGGCGCGGGGCGCCGGCGCGCGCGGCGCGCGGCGC-CGCGCGGCGCCGC 1383

Qy        603 GGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAG 662
Db       1384 GGGGNCGCCGCNCCGGCGCCGCGCGCGCCNCCGCGCNCGGNCCGCGGCGGNCGNCCGGCC 1443

Qy        663 ATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTG 722
Db       1444 GCCNCNCGGCNCGCGCGGCGCCGCGGCGCGCGCGCCGGCGCCNCGCCGCNCCNCNCCGCG 1503

Qy        723 CTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCC 782
Db       1504 CCCGCNNCNCNNCCGNNCCGCGCGCCNNCGCGNCCGCGNCNCGCNNCNCCGCGCGGCG 1563

Qy        783 TATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTC 842
Db       1564 GCGCGCCGGCGCCGCCGGCGCCCGNCCGCNCGGCGCNCCGCCCCGCGCGNCNCGCGCGNC 1623

Qy        843 GCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACG 902
Db       1624 GNNCGCGCGCGCGCCGCCGCCGCGCGCGCGGCGNCGGCCGGCCGCNCGCGCGCGCGCCNC 1683

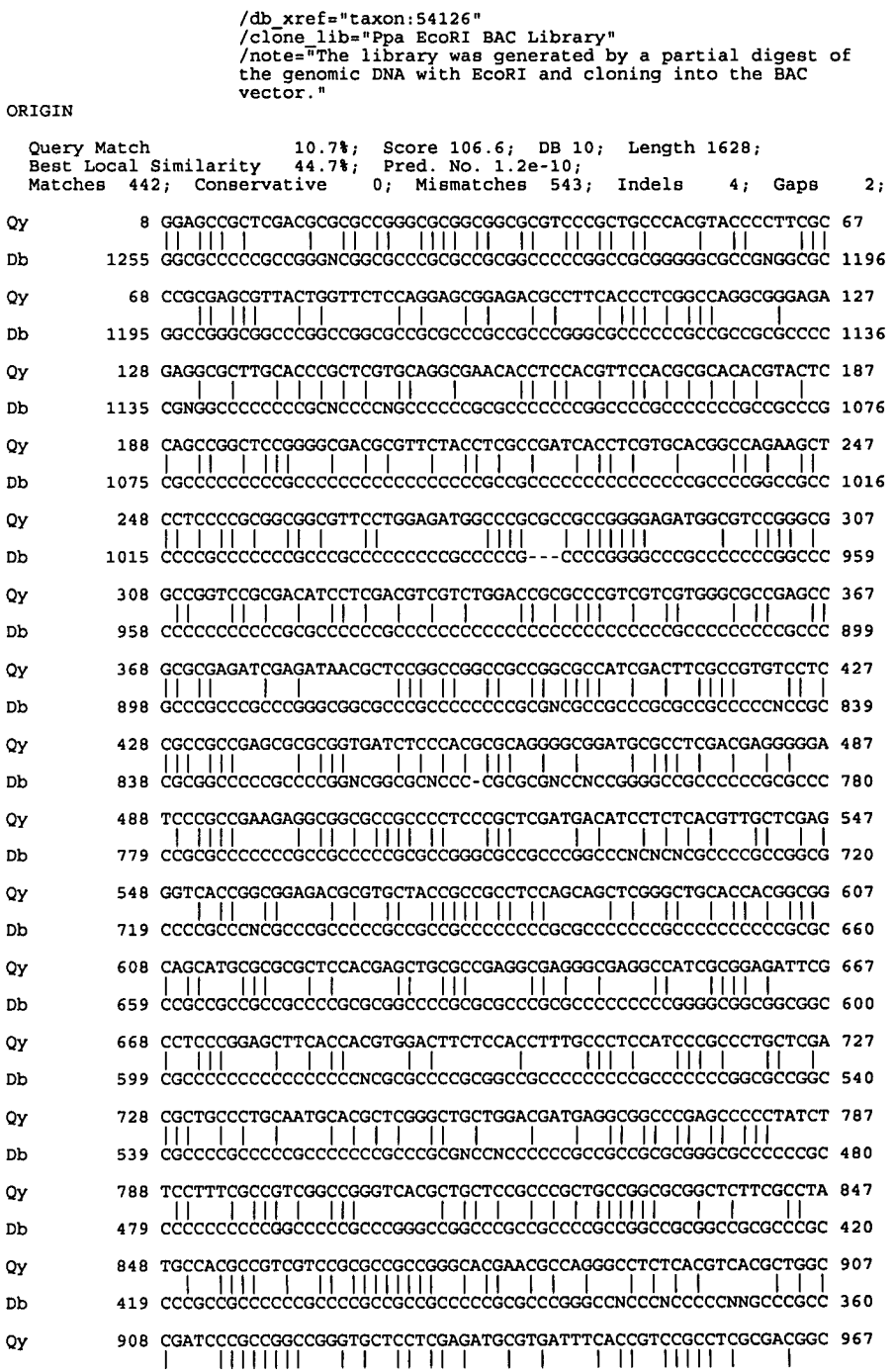
Qy        903 CTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCG 962
Db       1684 CGCGCCGCCGCCGCCCGNCGNCCCGGCNCCGCNNGGCGCCGCCNNCCCCCGGGCGCCGCG 1743

Qy        963 ACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTC 998
Db       1744 CCGCGCGNCGCGGCGCCCCGCCCGCCCCGGCCGCGC 1779
```

RESULT 3
CG757066/c

```
LOCUS       CG757066               1628 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P052-2-A02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG757066
VERSION     CG757066.1  GI:37985257
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1628)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1628
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN
```

Query Match 10.7%; Score 106.6; DB 10; Length 1628;
Best Local Similarity 44.7%; Pred. No. 1.2e-10;
Matches 442; Conservative 0; Mismatches 543; Indels 4; Gaps 2;



```
Qy          8 GGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGC 67
Db       1255 GGCGCCCCCGCCGGGNCGGCGCCCGCGCCGCGGCCCCCGGCCGCGGGGGCGCCGNGGCGC 1196

Qy         68 CCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGA 127
Db       1195 GGCCGGGCGGCCCGGCCGGCGCCGCGCCCGCCGCCCGGGCGCCCCCCGCCGCCGCGCCCC 1136

Qy        128 GAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTC 187
Db       1135 CGNGGCCCCCCCCGCNCCCCNGCCCCCCGCGCCCCCCCGGCCCCGCCCCCCCCGCCGCCCG 1076

Qy        188 CAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCT 247
Db       1075 CGCCCCCCCCCCGCCCCCCCCCCCCCCCCCCGCCGCCCCCCCCCCCCCCCCGCCCCGGCCGCC 1016

Qy        248 CCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCG 307
Db       1015 CCCCGCCCCCCCGCCCGCCCCCCCCCCGCCCCCG---CCCCGGGGCCCGCCCCCCCGGCCC 959

Qy        308 GCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCC 367
Db        958 CCCCCCCCCCCGCGCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCGCCC 899

Qy        368 GCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTC 427
Db        898 GCCCGCCCGCCCGGGCGGCGCCCGCCCCCCCCGCGNCGCCGCCCGCGCCGCCCCCNCCGC 839

Qy        428 CGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGA 487
Db        838 CGCGGCCCCCGCCCCGGNCGGCGCNCCC-CGCGCGNCCNCCGGGGCCGCCCCCCGCGCCC 780

Qy        488 TCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAG 547
Db        779 CCGCGCCCCCCCGCCGCCCCCGCGCCGGGCGCCGCCCGGCCCNCNCNCGCCCCGCCGGCG 720

Qy        548 GGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGG 607
Db        719 CCCCGCCCNCGCCCGCCCCCGCCGCCGCCCCCCCCGCGCCCCCCCGCCCCCCCCCCGCGC 660

Qy        608 CAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCG 667
Db        659 CCGCCGCCGCCGCCCCCGCGCGGCCCCGCGCGCCCGCGCCCCCCCCCGGGGCGGCGGCGGC 600

Qy        668 CCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGA 727
Db        599 CGCCCCCCCCCCCCCCCCNCGCGCCCCGCGGCCGCCCCCCCCCGCCCCCCCGGCGCCGGC 540

Qy        728 CGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCT 787
Db        539 CGCCCCGCCCCCGCCCCCCCGCCCGCGNCCNCCCCCCGCCGCCGCGCGGGCGCCCCCCGC 480

Qy        788 TCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTA 847
Db        479 CCCCCCCCCCGGCCCCCGCCCGGGCCGGCCCGCCGCCCCGCCGGCCGCGGCCGCGCCCGC 420

Qy        848 TGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGC 907
Db        419 CCCGCCGCCCCCCGCCCCGCCGCCGCCCCCGCGCCCGGGCCNCCCNCCCCCNNGCCCGCC 360

Qy        908 CGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGC 967
```

```
Db        359 CCGGGCCGCCGGCGCCGCGGCCCGCGCGCCCCCGGNCGCCNCCCCCCGCCGCCGCCCCCG 300

Qy        968 GGACGTCGCGCCCACCCCCGCCCAGCGGC 996
              |  |  ||| |||  || ||||  |||||
Db        299 GCGCCCCGCCCCCCGCCGCGCCGCGCGGC 271


RESULT 4
AG030607
LOCUS       AG030607               1538 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-003A18.F, genomic survey sequence.
ACCESSION   AG030607
VERSION     AG030607.1  GI:16557480
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1538)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1538
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-003A18.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             10.6%;  Score 106.4;  DB 10;  Length 1538;
  Best Local Similarity   42.3%;  Pred. No. 1.3e-10;
  Matches  390;  Conservative    0;  Mismatches  529;  Indels    2;  Gaps    2;

Qy         15 CTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCGCGAG 74
              | || | | |||       |||      |  || | ||     |||   | | | ||
Db        420 CGCGCCCCCCGCGCCCGCGCGCGCGCGCGCGGGCCGNCCCGCCCCCCGCGCCCGCCCGCC 479

Qy         75 CGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAGGCGC 134
              ||   | |   | |   | |||| | || |  | | | ||  |  || |  | | | |
Db        480 CGCGGCGGCGCCGCGGCGCGCGGCGNCGGCGCCGCGCCCGNGCGCGC-GCNGGGCGCCCG 538

Qy        135 TTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGCCGG 194
                ||  |  | || || |||      |  || ||  |      || | ||  | |  ||||
Db        539 CCGCGGCNCCGCGNGCCGGCCGNNGCGGCCCCGGGCGCGNGGCGCNCGCCCGCGCGCCGC 598

Qy        195 CTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCC 254
                 || | || |||       |  || |     | |  || | ||  |   | |  | |
Db        599 GCGCGCGCCGNCGCCCNGCGGCCGGCGGCCGCGCGCCCGCGCCGCGCGCGCCGCGGCGCG 658

Qy        255 GCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTC 314
              ||    | |  |  |      |||| ||||   || | |   |||  |   |||  |||
Db        659 GCCCGCGNGCGCGCGNCCGCGGCCNGCGCGCGCGCGCCGCCCGCGCCGCCGCGCGCGGCG 718

Qy        315 CGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAG 374
              ||    || || || || | ||  ||| ||   || ||     | |  ||  ||||
Db        719 NCCGNNCGCCCCGCCGCCGGCGGGCGCGCCCCGCGCGGCGGCNCAGNCNCGCGCCGCGGC 778

Qy        375 ATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCC 434
              ||      ||| |||  | || || |    |  | |  |||         || |||||
Db        779 GCCGNCGNGCCGCNCCGCCGCCGGCGGCGCGCGCGGCCGCCGCGCCCCGGGCCNCCGCC 838

Qy        435 GAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCC 494
              |||||||  |  | | |  ||   |  ||||  |    ||      | |  | |
Db        839 NCGCGCGCGCCGCGCGCGCNGGCCGCGCCGCGGCGCCCGGGCGGNCNCCGCGCCGCGCCG 898

Qy        495 GAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACC 554
              | |  | || |  ||| ||  |    |    |    |    |  |  | | | | |
Db        899 NGCGCGCGGNCGGCCGCCCGCCGNCGGNGCGCGGGCGCGCGCGGCCCCCNGNGCGCCGCG 958

Qy        555 GGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATG 614
              || | | || || |  | | ||        |  ||    || |    ||| | |   |
Db        959 CGCNGCGCCNCGCGCGAGNGNCNCCAAACCNCGGGCGCCGCGCGCGCGCGCCCGNNCG 1018

Qy        615 CGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCG 674
              |||||  |     |  ||  |  | | | || | |  |  |     |  | | | ||
Db       1019 NNCGCGCNCGCNCCGGCGCCNCCGCGGGCCGCCGCGCGCGCCNGCCCGCCCCNCNNCGCG 1078

Qy        675 GAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCC 734
                  | || ||    |    |      | |   | || || || ||   | ||
Db       1079 CGCGGNNCNCNCGCGGCNGCCCGNNCGCGNCNGCGCGGCCGCGGCCCGCGCGCGCCNGCG 1138

Qy        735 CTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTC 794
              |      ||  || |  || |        | |  | |||  ||| |        |
Db       1139 CCCGGCNGC-GGCGCNCGCCGGNCNCNNGCGCGCGGNCCCCNGCCNCGCGCGCCGCGNNG 1197

Qy        795 GCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACG 854
              ||   ||  | |    ||| |  |  | |||  || ||||| | |  |||    || |
Db       1198 GCGCGCGCTCCGCGNCCGCCGGCCGCCGCGCGNCCCGCGCGNCGCCCCGCGNNCGCGAAA 1257

Qy        855 CCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCC 914
              ||    |   ||   || || ||  |  |    || | | |  | |    ||    | |
Db       1258 ACGGAAGCNCGGCGCGCGNGCGCGNNCCGCGCNNCCGCGCNCCCCGCCGCCGCGNNCCNC 1317

Qy        915 GCCGGCCGGGTGCTCCTCGAG 935
              ||  |   | |  |  || |
Db       1318 CCCCNCGNNGCGNCCNCCGCG 1338


RESULT 5
AG393083/c
LOCUS       AG393083               1878 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-211E23.TJ, genomic survey
            sequence.
ACCESSION   AG393083
VERSION     AG393083.1  GI:48018931
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1878)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1878
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-211E23.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             10.5%;  Score 105;  DB 10;  Length 1878;
  Best Local Similarity   44.5%;  Pred. No. 2.4e-10;
  Matches  407;  Conservative    0;  Mismatches  501;  Indels    7;  Gaps    1;

Qy         13 CGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCGCG 72
Db       1634 CGCCGCGCGCGCCCCCGCCGCCGCGCCCGCGCGCGCGCGCCGCGNCCCGCCGCNCCGNCG 1575

Qy         73 AGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAGGC 132
Db       1574 CGCGCGGCGCGCCGCGCGCGCGCGCCCGCCGCGCNGCGCGCCGCGCCGCGCGCCGGCGCG 1515

Qy        133 GCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGCC 192
Db       1514 CGCGCGCCCGCNGCCCGCGGCCGCGCGCGNCGCCGCCCGCCGCGNCGCGCGCGCNGGNCC 1455

Qy        193 GGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCC 252
Db       1454 GCGCGCGCGCGCCGGCGCGCGCCGCCGCCGCGCCGCNCGCGNCCCCGCGCCCGCCGCCCGC 1395

Qy        253 CCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGG 312
Db       1394 GCCGCGCNCGCCGCGCGCGCNCGCCGCGNCCCGCCCGCCGCGCGCCCGCCGCGCGCCCGN 1335

Qy        313 TCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCCGTCGTCGTGGGCGCCGAGCCGCGCG 372
Db       1334 NCCGCGGCGCGCGCGCGGCGCCCCGCGCCCCCCCGCGGCGCGCCGCCGCCGCGCGCGCC 1275

Qy        373 AGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCG 432
Db       1274 GNGCCCCGCCGCGNCCGCGCCGGCGCGGGGCCGCCCGCCCGCGCCCGCGCCGCGCCGCGG 1215

Qy        433 CCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCG 492
Db       1214 CCGCGCCCCGCGCGCGCCCGCCGCCGCCGCCGCGCGCGCCCGGCCGCGCGGCNGNGCCCG 1155

Qy        493 CCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCA 552
Db       1154 CCGCGCCGCGCGCCCCGCGNCCGCGCCGCCCCCCCGCCGCCGCGCCCGGCCCGCCGCCCC 1095

Qy        553 CCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGC-------TGCACCACGGC 605
```

```
Db       1094 GCCGCCNCGCGCGCCGCGCCCGCGCGCCCCCGCCGGCCGNCCGCCGCCGCGCCCGCCGGC 1035

Qy        606 GGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATT 665
Db       1034 GGNCCCGCGGCCGCCGCCCGCGCGCNGGCCCGCCGCCGCCGCGCGCCCCCGGCGCCCCCC 975

Qy        666 CGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTC 725
Db        974 GCCCGCGCCCGCCGCCGCCGCCGCGCGCCCCGCGCCGCGCGCGGCCGCGCGCCCCCGCCCC 915

Qy        726 GACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTAT 785
Db        914 GCCGCCGCCCCGCCGCCGCGGCGCGGCCGCGCGGCCCGCCCGCGCCCGCCCCGCCGCCCC 855

Qy        786 CTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCC 845
Db        854 CGCGCCCCGCNCCCGCCCGGCGCGGCCGCGCCCCGCCGCGCGCCGCCGCCGCGGCCGC 795

Qy        846 TATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTG 905
Db        794 CGCGCCCCGCCCCCGCCCGCCCCGCGCCGCCCCCGCCCCCCGCCCCGCCGGCCGCCCCG 735

Qy        906 GCCGATCCCGCCGGC 920
Db        734 CCCCGCCCCGCCGCC 720
```

```
RESULT 6
AG030608/c
LOCUS       AG030608               1189 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-003A18.R, genomic survey sequence.
ACCESSION   AG030608
VERSION     AG030608.1  GI:16557481
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1189)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1189
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-003A18.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
```

```
  Query Match             10.3%;  Score 102.8;  DB 10;  Length 1189;
  Best Local Similarity   46.7%;  Pred. No. 6.4e-10;
  Matches  361;  Conservative    0;  Mismatches  400;  Indels   12;  Gaps    2;

Qy        158 CACCTCCACGTTCCACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCT 217
              | |   | ||  || |||||   ||      |   || |  || ||||
Db       1147 CGCNCGCCCGCGCCGCGCGCCGCNCGCGGCGCGCGCGCCCCGCGCGGCGCGCGCGGGGCGGG 1088

Qy        218 CGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGC 277
              || ||  | |  || ||  | |        |  | ||       ||  |  || |   ||
Db       1087 CGGCGCGCGCGGCGGGCGGGCGCGCGGCGCGCGGCGCGNCCGCCCGCGCGNGGGGGGNGC 1028

Qy        278 CCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTG 337
              | |||  |  | | |  || | | |||||  |   ||| |  || || || || |
Db       1027 GCCCGCGNCGCGCGCGGCGGGGCCGGCGCGGCGCGGGGGCGGCNGCCGCG-CGGCGGCGC 969

Qy        338 GACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGC 397
              |  |||| |||     | | |    |    |||||  |   | |     ||| | |  ||||
Db        968 GGGCGCGGCCGGGCCGGCGNGGCGGGGCGGGCGCGCGGCGGCGGCGGCGCGCGGCGCGGC 909

Qy        398 CGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGC 457
              || ||||||  |  |  |||   | |    | || |   |||||     | | | ||
Db        908 GGCGGGCGCCCGCCGCCCCGCGNCGNGCCGGCCGGCGGGCGGCGCGCCCGGCGCGCGCGN 849

Qy        458 GCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCC 517
              ||  |||||    || | || || | | |  | ||  |  | |  |||||    | |  |
Db        848 GCGCGGGCGCGGCCGGCGCGCCGCGCGCGGGCGCGGGGCGGCGCGGGCGCGCGCGCNGCG 789

Qy        518 GCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCG 577
              |   | | |              |  ||  | ||  | |||  | | |      ||  ||
Db        788 GGCGGCGGGCCGGGGGGGCGGGCGCGGGCGGCGCGCCGCGGCGGGCCGGGCGNCCCCGCG 729

Qy        578 CCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCG 637
              ||  | || |  || ||  | || ||||||| ||  |   ||     | || ||  || ||
Db        728 CCCGCGGCCGGGCGCGCGCCCCCCCGGCGGCGGCCCGNCGGCCGCGCCCGCGCGCCGGCG 669

Qy        638 AGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTC 697
                ||  | ||  |||  ||||  |      |   |||     | |   | |        |
Db        668 GCCCGCCGCCGGCGCCCGCGCGCCGGCCGCGCGGNCGGCCGCGCGCNCGGCGCGNGCGGC 609

Qy        698 CACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCT 757
              |  |   || | ||  | ||| | || || ||| || | |   |    | ||| |  | ||
Db        608 CCGCGCCGCGCCCCGCCGCGCGCCGCGCGCCGCCGCGCGNCGCCCCGCGCGNGCCCGGC- 550

Qy        758 GGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCT 817
                      || |  || || | |    |  ||   ||||| |||||||       |
Db        549 ----------GCCGGGCGGGCGCGCGGGCCGCCGCGCGCCGCCGGCCGGCCGGGCCGCGC 500

Qy        818 CCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCAC 877
              | || | |   | |||  || |  || |   ||    |||  | ||||| | | |
Db        499 CGGCGCCCGCGCCGCGGCGCCCGCCGGCGGCGCGCGCCCCGGGGGCCGCGNCCGCCGCGCG 440

Qy        878 GAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCC 930
              |  |||  || | |  | ||   ||| ||  |  | ||| |||  ||  | ||
Db        439 GCGCGCGCGGCCGCCGCGCGCGCCGCCGGNGGCGCGCGCGGGCGNGGCNCGCC 387
```

```
RESULT 7
AG430101/c
LOCUS       AG430101               1552 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-304P19.T7, genomic survey
            sequence.
ACCESSION   AG430101
VERSION     AG430101.1  GI:48073164
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1552)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1552
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-304P19.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

```
  Query Match             10.1%;  Score 100.6;  DB 10;  Length 1552;
  Best Local Similarity   44.5%;  Pred. No. 1.7e-09;
  Matches  426;  Conservative    0;  Mismatches  531;  Indels    1;  Gaps    1;

Qy         39 CGCGTCCCGCTGCCCACGTACCCCTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGA 98
              ||||  |||| ||   ||  |||       |  |  |       |  ||  ||  | |
Db       1549 CGCGCGCCGCCGCGGCCGCGCCCGCCGCGCGCCCCGGCCCGCCCCCCCGCCGCCCGCCGC 1490

Qy         99 GACGCCTTCACCCTCGGCCAGGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAAC 158
                | |  | ||| ||||| | | |  | |  ||||     | |      || ||||
Db       1489 CCGGGCGGCGCCCCCGGCCGGCCCGCGGGGGCGCGCGCCGCCGCCGCGCCGCGGGCGCCG 1430

Qy        159 ACCTCCACGTTCCACGCGCACACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTC 218
              || || ||    | | | | | |   | |  |||| | | | ||||  |      |  || |
Db       1429 GCCGCCCCGCGGCCCCCCCGCCCCGCCGCGCGCCGCCCCGGCGGCGGCCGCCCCGCCCGC 1370

Qy        219 GCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCC 278
              ||  |  |     | |||||   || ||  |     ||  ||  || |     ||
Db       1369 CGCGCGCGGCGGCGCCCCGGCCGCGCGCGCCCGCGGCGCGCCCCGCCCCCGCCCGGCGCG 1310

Qy        279 CGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGG 338
              |||||||||   | |   | | |  |||||||| |||||  |  || || |  || | |
Db       1309 CGCGCCGCCCCCGCGCGCG-GCCGCCGCGGCCGGCCCGCGCCCGCCGCGCCNCCGGCGCG 1251

Qy        339 ACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCC 398
               || ||| ||    |  |  |||||     |||| |  |  |      |  | |||||
Db       1250 CCCCCGCGCGCGCCCCCGCCCGCCGCGGGCGCCCGGCCCCGGCCCGGGCCGGGCCGCCG 1191

Qy        399 GCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCG 458
              |||||   |  |       ||  | || || |||||  ||||  ||          || |||
Db       1190 GCCGGGCGCCCCCGGGGGCCCCCGCCCCCCCCCGCCCCGCGCCCGCGCCCGGGCCGCGCC 1131

Qy        459 CAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCG 518
              |  ||       | |      || |  |   || ||||    |||  |||||| ||  ||
```

```
Db       1130 CGCGGCGCCCGCCCCGCGCCCGCGCCGCCGCCGGCCGCGCCGGCCCCGCCGCGCCCGCCC 1071

Qy        519 CTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGC 578

Db       1070 GCGCCGCCCCCGCGCCGGCGCGCCCGCGGGCCCCCGCCGGGCGGCCCCCGCCGGCCGGCC 1011

Qy        579 CTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGA 638

Db       1010 CCCGGGGCGGCGCCCCCCCCGCGCGGCCGGCCCCCCCCGCGGCGCCCGGGCCCCCCGG 951

Qy        639 GGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCC 698

Db        950 GGGCGGGGCGGCGCGCCCCCGGCCCGGCCCCCCGGCCGCGCCGCCGCCCCGGCCGCC 891

Qy        699 ACCTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTG 758

Db        890 GCCCCGCGCCCCGGCNGGCCGCCGCCCGCCCCGGCCGCCCCGCCCGCGGGCCCCGGGCGG 831

Qy        759 GACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTC 818

Db        830 GCCCCCCCCCCCGCGGGCCCCCCCCCCGCCGCGGGGGCCCGGCGCCGGCCCCGGCCGGGG 771

Qy        819 CGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACG 878

Db        770 CGCGCCCGGGCCCGGCCCCGCCCCGCGCGGCNCGCGCCCCCCGCCCGCCCGCCGGGGNCG 711

Qy        879 AACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATG 938

Db        710 CGGCNCCGCCCCCGCGCCCGCCGGCGCCGCCCCGCCGCCCGCGGGGCGCCCCCCCCCGCC 651

Qy        939 CGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGC 996

Db        650 GCGGGGGGCGCCGCCCCNNCCGGGGGCGGCCCCCCCGCCCGCCCCGCGGCGGGCCCGC 593
```

RESULT 8
AG435522/c

```
LOCUS       AG435522               1961 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-313K12.TJ, genomic survey
            sequence.
ACCESSION   AG435522
VERSION     AG435522.1  GI:48078585
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1961)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1961
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-313K12.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

```
  Query Match             10.0%;  Score 99.8;  DB 10;  Length 1961;
  Best Local Similarity   44.7%;  Pred. No. 2.4e-09;
  Matches  443;  Conservative    0;  Mismatches  529;  Indels   19;  Gaps    2;

Qy          8 GGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGC 67

Db       1780 GGCGCGCCCGNCCGCCCCCCGCCCCCGCCCGGCGCCCGCGCNGCGCCCGCCCCCGCCCCG 1721

Qy         68 CCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGA 127

Db       1720 CGCCGNCCCCCCCCGCGCCCCGCCGCCCGCGCCCGCCGCCCCCCGCGNCCCCNGCCGCCC 1661

Qy        128 GAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTC 187

Db       1660 CCGGCCCCGCCCCGCCCGGCCGCGCCGCCCCCCCCCCGCGCCCCNCCCGCGCCGCCGCCG 1601

Qy        188 CAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCT 247

Db       1600 CCGCCCGCGCGCGCCCCGCGCCCCCGCCGCNCCCGCGCCCCCCGCCCGCCGCCCCGCCCG 1541

Qy        248 CCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCG 307

Db       1540 CCCNCCGCCCGCCGCCCCCGCCNCCCCCGCGCCCCCCCCCGCCGCCCCCCCGCCGCCGCC 1481

Qy        308 GCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCC 367

Db       1480 NCCGGCCCGCGCGCCCGCCCCGCCGCGGCCCCCCCCCCCCCCCCNGCGCGCNCCCCCCGGC 1421

Qy        368 GCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTC 427

Db       1420 GCCGCNCCCCCCGCGCNCCCGCCCGCCCCGCGCCCGCCCCCGGCNCGGCGNCCCGCCCCC 1361

Qy        428 CGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGA 487

Db       1360 CGCCGCNCCCCCCCCGCG-----CCCCGCCCCCCGGCGCCGCCGCCCCCCGGCCCCCGCCC 1306

Qy        488 TCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAG 547

Db       1305 GCNCCCCNGCCCGCCCCCCCCGCCCCCCCCCCGCGCCCCCGCCCCCCCCCG--------- 1255

Qy        548 GGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGG 607

Db       1254 -----CCCGCGCCCCCCCCCGCCGCCGCCCCCGCCCCCCGCCGCCCCCGCCCCCCGCGC 1200

Qy        608 CAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCG 667

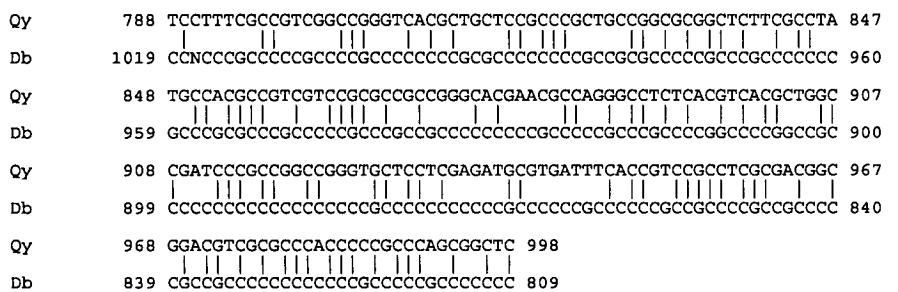
Db       1199 CCCCCCCGCCCCCCCCCCCCCGNGCCCCCCCCGCCCCGCCCCGCCCCCCCCCCGCCCCGCC 1140

Qy        668 CCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCTCGA 727

Db       1139 CCGCCCGCCCCGCCCCCCCCCGCCCCCGNCCCCCCCGCCCGCCCGCCCCCCCCCGCCGCCC 1080

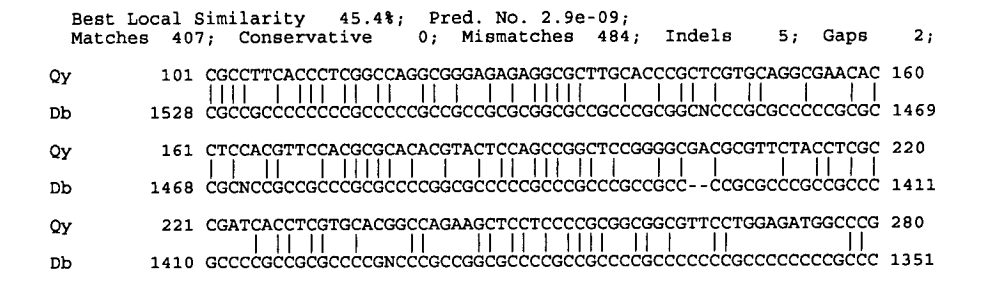
Qy        728 CGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCT 787

Db       1079 CCCCCCCCGCCCCCCCCCCCGCCCCCCCCGGCCCGCCCCCCCCCCCCCGCCCCCCCCCCC 1020
```



RESULT 9
AG430426/c
LOCUS AG430426 1531 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-305H10.T7, genomic survey sequence.
ACCESSION AG430426
VERSION AG430426.1 GI:48073489
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1531)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1531
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-305H10.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 9.9%; Score 99.4; DB 10; Length 1531;



RESULT 10
AG030611/c
LOCUS AG030611 1341 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-003A20.F, genomic survey sequence.
ACCESSION AG030611
VERSION AG030611.1 GI:16557484
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1341)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1341
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-003A20.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             9.9%;  Score 99;  DB 10;  Length 1341;
  Best Local Similarity   45.7%;  Pred. No. 3.5e-09;
  Matches  454;  Conservative    0;  Mismatches  515;  Indels   25;  Gaps    6;

Qy          11 GCCGCTCGACGCGCGCCGGGCCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCG 70
               |||||    |||||||  | ||||  |  |||||  |    || | ||  | |   ||||||
Db        1099 GCCGCCGCGCGCGCGCNGCGCGCNGGCGCGCGCGCGCGCGCGCGCCCGCGCCCGCCCG 1040

Qy          71 CGAGCGTTACTGGTTCTCCAGGAGCGGAGA---CGCCTTCACCCTCGGCCAGGCGGGAGA 127
               ||  |     | |   | |      |||  |    |||    | | | ||  |  ||| | |
Db        1039 CGCCCNCCGCNGCGGCGCNCNCCGCGCCGNNCCCGCGCGCGCGCCCGCGCGCGCGCGCGC 980

Qy         128 GAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTC 187
               | |||||  |   |||| || |  | ||  | || || ||   | ||||| | |     |
Db         979 GCGGCGCGNGGGGCCGCGCGCGGCGNCGCGCGCCGCCGCGNCGCCCGCGCCCNCCGCGCC 920

Qy         188 CAGCCGGCTCCGGGGCGACG--CGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAG 245
                  || || |||  ||| ||  ||       | |||||              ||  | |
Db         919 GCCCCCGCCCCGCCGCGGCGCCCGGCGCGNCCCGCCGCCGCGNGNNGCGCCGCGCCGCGC 860

Qy         246 CTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGG 305
                  | | |     |||||         |  | | | | |   ||  |   ||     |
Db         859 GCGCNCGCGCGNNCGGCGCGGGGNNGGCGGNGCNGNGNGGNGNGGNNGCGCGCCCGNCGC 800

Qy         306 CGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAG 365
               || ||   ||        |  ||    |  | |  |||||| ||  | || | ||   | |
Db         799 CGCCCCCGCCCGNCGCGCGCCGCGCGCCGCNGCCCCGCGCGCGNNGGCGCGCGCNNGGCG 740

Qy         366 CCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCC 425
               ||||||              ||| |   ||| |||||| |  |  |  |  |||||    |
Db         739 CCGCGC---------CCCGCGCCCGCCCCGCCCGCCGCCCGCCGCCGCNNCGCCGCNCNC 689

Qy         426 TCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGG 485
                 |  |  ||  |||||| |     | | | | |||   | |||  ||||  || ||
Db         688 CCGCCGNCCCCGCGCGCCGCCGCCGCGCGCCCGCCCCGCCGGCCGCGCGGCGCCG----- 634
```

```
Qy         486 GATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCG 545
               | ||| |  | |||| ||| | ||  |||| |   | |        | ||  || | |
Db         633 --GCGCGCGGCGGCGGCGNCGCGCCGCCCGCCGCGCCNCGCCGCNNGGGCGCGGCGCGCC 576

Qy         546 AGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGC 605
               |  | | |   |  |||  ||   ||  ||| |  |  || | ||  ||  |   ||
Db         575 GCGCCCGCNGCGCGCCGCGCCCGCGCGCGGGGCCGCGGGGCGCGCCGGGCGCGNCGGCGC 516

Qy         606 GGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATT 665
               ||| ||  |  |||| || ||  |   ||| |||| | | | ||    |||||  |
Db         515 GGCGGCGCGGCGGCGCCCCGCGGCCGCGCCCGCGGCGCGCGGGGGGGGCCCGCGGGGCGC 456

Qy         666 CGCCTCCCGGAGCTTCAC-CACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGCT 724
               |  |||| ||  | | | ||  |   |  | ||   || | |   |  ||| ||
Db         455 GCCGCCCCGCCGCGCCCCGCGCGCCGCGCGCCGCGCCGCCGCGCGCGCCGCGCCCCGGCC 396

Qy         725 CGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTA 784
               || |||  ||  ||       | | || ||  |      |  || | ||   |||| |
Db         395 CGCCGCCCCCGCGCCCCCGCCCCCCGCGCCCCCCCGGGGCCGGGGGCCCGCCGCCCGCGC 336

Qy         785 TCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTC---CGCCCGCTGCCGGCGCGGCTCTT 841
               |  ||   | ||  |    |   |  ||   |  | | ||| || |   |
Db         335 CCCCCCCACAGGCGAAGCAATCTTTTTTTTTTTCTTTGTGCCCCGCCCGCCCCCGCCCC 276

Qy         842 CGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCAC 901
               ||   ||  |||    |||| ||  | ||    | | | | | | | | | | | | |
Db         275 GCCCCGCCCCCNGCCCCGCCCCGCCCCCGNGNGCNGCGCCCCCCCGCCGCCGCGCCGCGC 216

Qy         902 GCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGC 961
               | |   |   |  |||| |     |   |  ||  |       |  || ||
Db         215 GNNGCGTGNNGGCCGNGGCCCGCCCGCGCCNCCGCGGCCCGGNCNNNNCNCCCCCCCNCG 156

Qy         962 GACGGCGGACGTCGCGCCCACCCCCGCCCAGCGG 995
               | |   |  | | ||| ||||| ||| | ||
Db         155 CCGGCCNNGCCCCCCCCCCCCCCCCCCCCGGGGG 122
```

```
RESULT 11
CC210735
LOCUS       CC210735               1159 bp    DNA     linear   GSS 12-MAY-2003
DEFINITION  CH261-185P7_Sp6.2 CH261 Gallus gallus genomic clone CH261-185P7,
            genomic survey sequence.
ACCESSION   CC210735
VERSION     CC210735.1  GI:30529403
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1159)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 71
            High quality sequence stop: 159.
FEATURES             Location/Qualifiers
     source          1. .1159
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-185P7"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
```

```
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             9.8%;  Score 98.2;  DB 9;  Length 1159;
  Best Local Similarity   45.6%;  Pred. No. 5e-09;
  Matches  393;  Conservative    0;  Mismatches  466;  Indels    3;  Gaps    2;

Qy        138 CACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAGCCGGCTC 197
              | ||| | |   |  |     | || || |    || |||   | |      || |||  | |
Db        234 CCCCCCCNCNCNCCCGGCCCCGCCCCCCCCCCCCCCCGCNGCCGCCCCGCCCGGCCCCCGC 293

Qy        198 CGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCTCCCCGCG 257
              | |  || ||||      || |  ||  | |  ||  | | |||      | |  ||||||
Db        294 CCGCCCGCCGCGNCGCCCCCCCGCGCCCGCGGCGCCCGCNGCCCCCCCCCCGCCCCCGCC 353

Qy        258 GCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCCGGTCCGC 317
                |||   ||    |  | || || |||||  |    | || ||   |  || | || |
Db        354 CGGGCCCGCCCCCGGCCGCCCGGCCCCGCCNCGCGCCGGCCGGCCC--CCCCCGCCCCC 411

Qy        318 GACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCGCGAGATC 377
              | |  ||    ||  |       |||| |||| |  |   ||||    ||| | |     |
Db        412 GCCCNCCCGCCCGCGGGGCCCCCCGCCCCCGCCCGCCGCGGCGGGCCGCCCCCCCCCCC 471

Qy        378 GAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGCCGCCGAG 437
              |      | | || | ||  ||||  | ||  || |  | | |   ||  ||||||||
Db        472 GCCCGGGCCCCCCCGGCGCGCGCCCCCCCCGCCGCCCCCCCGGCCCCCCGCGCCGCCGGC 531

Qy        438 CGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCCCGCCGAA 497
              | | |   |  | | | | | |              || || |  ||  |       |||||
Db        532 CCCCCCCGGGCCCCGCCCCCCCCCCCCCGCCCCGCCCCCCCCCGCCGCCCCGGCCGCCCCC 591

Qy        498 GAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACGTTGCTCGAGGGTCACCGGC 557
              | | |  |||| |||| | |||  ||  | |  || |         | |  |       |
Db        592 GGGCCCCCGCCCCCCCGCGCGCCCGCCGCCCNCCCCGGCGCCCCCCCCCGCCCGCGCCCG 651

Qy        558 GGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACGGCGGCAGCATGCGC 617
              |    ||||  |   | ||  |  | ||   | | |  | || | |  | ||    | |
Db        652 GCCCCCGCGCCCCCGCCCCCGCCGCCGCCCGGCCGCCCCCCCCCCCCCCCCGGCCGCCCC 711

Qy        618 GCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGATTCGCCTCCCGGAG 677
              || | ||  |  ||   |||  ||       |   | | |       ||||  ||||
Db        712 GCCCCCCCCCGGCCCGCCCGCCGCCCCCCCCCCCCGCCCCCCCGCCCCCGCCNGCCGGGC 771

Qy        678 CTTCACCACGTGGACTTCTCCAC-CTTTGCCCTCCATCCCGCCCTGCTCGACGCTGCCCT 736
              |  | ||        |  | ||  |    |||| ||  ||| | |  |  | |  | || |
Db        772 CCGCCCCGNCCCCCCCCCGCCCCGGCCCGCCCCCCCCCCCCCGCCCCCCCCCCCCGCGCC 831

Qy        737 GCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTCCTTTCGC 796
              ||    | | |  ||           |    |||| | ||||   |  ||   |||
Db        832 GCCGCCCCCCCCCCCGCCCNCCCCGCCCGGCCGCGCCCGCGGCCCCGCCCGCCCCCCGC 891

Qy        797 CGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATGCCACGCC 856
              |  || ||    | | |  |  | ||| |  ||       | |  ||||    || ||||
Db        892 CCGCGCCCCCCCCCCNCCCCCGCCCCCCCCCCCCCCGCGCCCCCGCCGCCCGCCCCCCGC 951

Qy        857 GTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCGATCCCGC 916
              |  || | |||  || |    || | ||  || | | ||  ||   |||  |   |
Db        952 CCCCCCCCCCCGGCCCCGCCCCCGCCCCCCCCCCCCCCCCCGCCGGGCCCCCCGCCCGGCC 1011

Qy        917 CGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGGACGTCGC 976
              | |||    | |  |  |  ||      | ||  ||||| |||| |  |   || |
Db       1012 CCGCCCCCCCCCCGCCCGGGCGCCCCGCCCCGCCCCCCGCCCCGCGCCCCCCCCCGCCCG 1071

Qy        977 GCCCACCCCCGCCCAGCGGCTC 998
              | |  ||||  |||  | || |
Db       1072 GGCGCCCCCGCCCCCCCCGCCC 1093


RESULT 12
CL078342
```

```
LOCUS       CL078342               1674 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-149D22_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-149D22, genomic survey sequence.
ACCESSION   CL078342
VERSION     CL078342.1  GI:40534255
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1674)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 300
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1. .1674
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-149D22"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             9.8%;  Score 98.2;  DB 10;  Length 1674;
  Best Local Similarity   45.2%;  Pred. No. 4.9e-09;
  Matches  445;  Conservative    0;  Mismatches  527;  Indels   13;  Gaps    2;

Qy         11 GCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCG 70
              ||| | |  ||| | || | | |  | ||     |||||  ||| |     | |  | |||
Db        601 GCCCCCCCCCGCCCCCCCGCCCCCCCCGCCGCCCCCGCCCCCCGCCCCGCGCCGCCCCCC 660

Qy         71 CGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAG 130
              |   ||    |     | |     | |   | |  | |||  |   ||    ||   | |
Db        661 CCCCCGCCCCCCCCCCCGCGCCCCCCGCCCCCGCCGCCCCCCCCCCCCCCCGCCCGCGCC 720

Qy        131 GCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAG 190
              ||||  || |||   ||  |       | || || |   || ||| | |     | |
Db        721 GCGCGCGCGCCCCGCCGCCCCCCGCCCCCCCCCCCCCCCGCCCCGCCCCCGCGCC 780

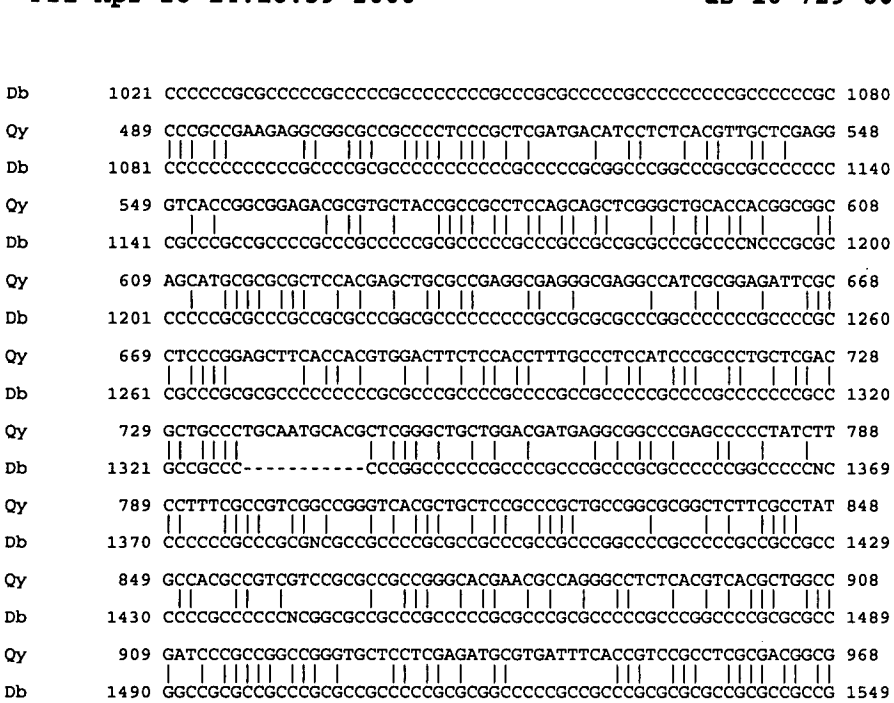
Qy        191 CCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCG--TGCACGGCCAGAAGCTC 248
              ||| |  || | |  | |    |  || |||||  | || ||    | || ||    || |
Db        781 CCGCCCGCGCGCCCCCCCCCCCGCCCCCGCCGCCCCCCCCGCCCCCCCGCCCCCCGCCC 840

Qy        249 CTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGG 308
              |  ||| |  |  ||   |       | | |  |||||  |       |  |    |
Db        841 CGGCCCCCCCCCCCGCCGCCCCCCGCGCCGCCGCCCGCCCCGCCCGCCCCCGCGCCCCCC 900

Qy        309 CCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCG 368
              |||  || |  |  || |     |  |    || |||||  |  |     | ||| |||
Db        901 CCGCCCCCCCCCCCCCGCCCGCCCCCCCCCCCCCCCGCCCCCCCGCCCCCCCCCGCGCCC 960

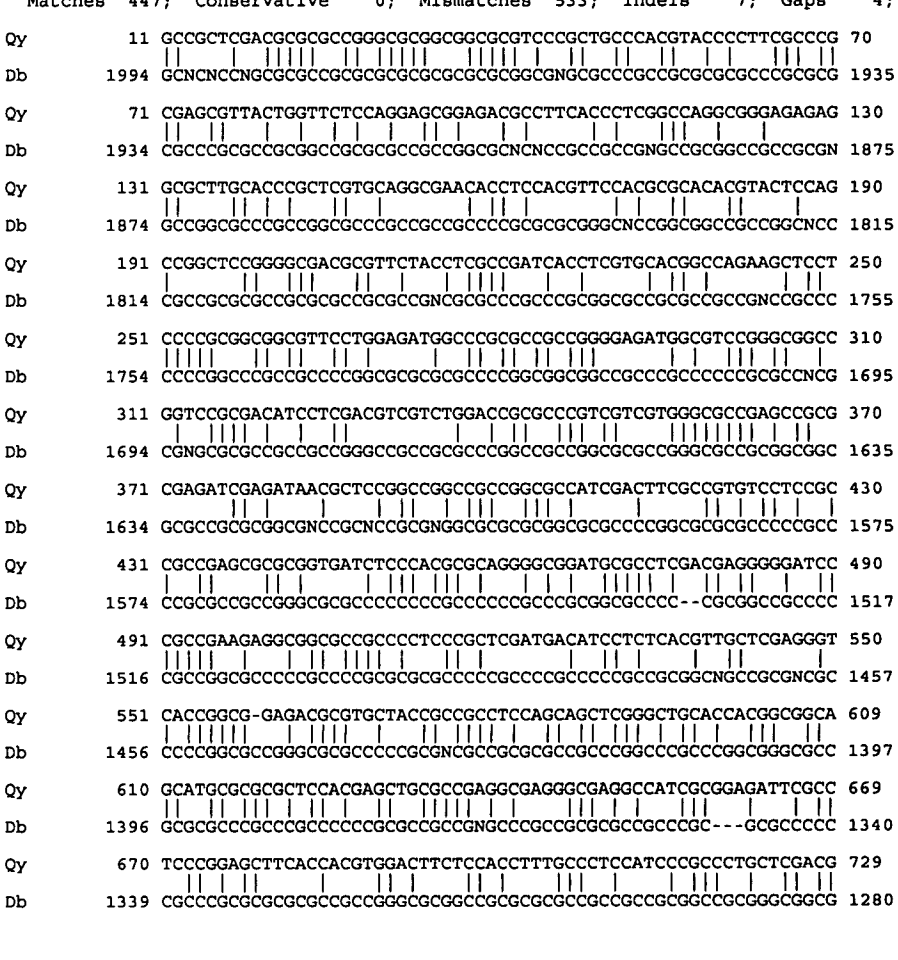
Qy        369 CGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCC 428
              | || |  ||          | |||  |  || || |   |  |     ||   || |
Db        961 CCCGCGCCCGCCCCGCGCGCGCCGCCCCCGCCCCGCCCGCCCCCCGCGCCCCCCCCCCG 1020

Qy        429 GCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGAT 488
               || ||| || | ||      ||| | |||  | ||    || || |  ||
```



RESULT 13
AG381986/c
LOCUS AG381986 2243 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-192D21.TJ, genomic survey sequence.
ACCESSION AG381986
VERSION AG381986.1 GI:47993191
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 2243)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .2243
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-192D21.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 9.8%; Score 97.6; DB 10; Length 2243;
Best Local Similarity 45.3%; Pred. No. 6.4e-09;
Matches 447; Conservative 0; Mismatches 533; Indels 7; Gaps 4;



```
Qy        730 CTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCTATCTTC 789
              | | |  ||      ||| || ||  |    ||  | |    |||| | |||      |
Db       1279 CGGGCGCGCNCCCNCCGCGCGCGCGCCGCNCCGGCGCGCNCCCCCCGCGGCCCGGGCGGGC 1220

Qy        790 CTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCGCCTATG 849
                |  |   |  | | |     | || |  ||  | ||||||  |  | |     |   |
Db       1219 GCCGCCCGCGCCCCGGCGCGCNCCCGCGCGCCCGCCCGCCGGCCCCCCGCGCGCGCGCCG 1160

Qy        850 CCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGCTGGCCG 909
              ||    |  ||  ||   | ||| ||     ||||   |   | | | | | |   |
Db       1159 CCCGCGCCCCGCNCGGCGCCCCGCGCGGCNGCGCCGCNGGGGCGCGGGCCCC-CGCCCGC 1101

Qy        910 ATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGACGGCGG 969
                ||| | |  |||  || |     |  |||      | |   ||||  | || |    |
Db       1100 CGCCCCCGGCGCGGCGGCCCGCGCGGGCGCGCCGCGGCGCGCCCCGCGGCCCGGCCCGCG 1041

Qy        970 ACGTCGCGCCCACCCCCGCCCAGCGGC 996
              ||  |||| | || |||||  ||| |
Db       1040 CCGCGGCGCGCGCCGCCGCCGCGCGCC 1014


RESULT 14
AG441877/c
LOCUS       AG441877               1462 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-323H09.TJ, genomic survey
            sequence.
ACCESSION   AG441877
VERSION     AG441877.1  GI:48084940
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1462)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1462
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-323H09.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             9.7%;  Score 97.2;  DB 10;  Length 1462;
  Best Local Similarity   46.0%;  Pred. No. 7.7e-09;
  Matches  461;  Conservative    0;  Mismatches  529;  Indels   12;  Gaps    5;

Qy          2 GGACTGGGAGCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCC 61
              || | | |   ||  || |||||||  ||  ||  |||||    ||| || | ||  |||
Db       1327 GGGCGGCGCCGCGGCCGGCGCGCGCGCGGGCCGCGGGCGCCGGGCGCGGCGCCCGGCCCC 1268

Qy         62 CTTCGCCCGCGAGCGTTACTGGTTCTCCAGGAGCGGAGAC---GCCTTCACCCTCGGCCA 118
              |  ||| || | |      | |   |  |  | || | | |    ||  | |  ||  |
Db       1267 CGCCGCGCGGGCGGCCGNCCGCCGCGGCCCGGGCCGCGCCCGGCCCGGCGCGGGCGCGCC 1208

Qy        119 GGCGGGAGAGAGGCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCA 178
              | ||| |   ||  |       || | |  || |||    | |    | ||   | |||||
Db       1207 GCGGGGGGCCGGGGCCCCCGCGCCCCGCCGGCCGGCCGGCGCGCGCGCCGCGCGCGCG 1148

Qy        179 CACGTACTCCAGCCGGCTCCGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGG 238
              | |      || ||  || || |  |||    |    |||  |  | || ||  ||   |
Db       1147 CGCCGCGCCGGGCGGGGCCCCGGCCCNCGCCGCCCCGGCCGCGGCCCGCCGCGCGCCGCG 1088

Qy        239 CCAGAAGCTCCTCCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGC 298
              ||    |  | ||   |  |||   | ||    ||||  |  ||||   |    |||
Db       1087 CCGCCGGGCGCGGCCGCGGCGCGCGCGNCGGGCCCCGGCCGCCCGCGCCCCCGNCCGGGC 1028

Qy        299 GTCCGGGCGGCCGGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGG 358
              | | || ||| | || |   | || || |     |  ||  | ||| || ||
Db       1027 GCGCCCCCGCCCG--CGGCCGCGCGCGGCGCCCCGGCGGGCGGCCGCGGCGGCGCCCC 970

Qy        359 CGCCGAGCCGCGCGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGC 418
              || |||| |||              || ||||| |||||   ||    || |  | | |
Db       969 GCGCGCGCCGGGCGCCCCGCCCCCCCGGCCCGGCGGGCCGGGCGCCGGGCCGCCGCCCC 910

Qy        419 CGTGTCCTCCGCCGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGA 478
              || | |  || |||||  ||  ||||      ||  ||| | |         |||   |
Db       909 CGGGGCGCCCCCCGCCCGGCCGGCGGGCCCGCCCGCCGCCCCGCCCGCCGGCGCGCGGC 850

Qy        479 CGAGGGGGATCCCGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCACG 538
              || |  |   |||||  | ||| |||| | ||| | ||  |  | | |  | | | | |
Db       849 CGGCGCCGCGGGCGCCG--GCGGCCGCGCGGGCCCGCGCGGCCCCCGGCGCCGGCCCCCC 792

Qy        539 TTGCTCGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCA 598
                |  |    ||||| |   |||| |    ||   ||  | ||  |   |||
Db       791 GCCGCCCCCGCCGGCCGGCCGCCGCGCGCGGCCGCGGGCCCGGCGGCCCCCGCGGCGCGG 732

Qy        599 CCACGGCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGC 658
              |   || ||  |  || |||| |    || |||||| ||| || |  | ||  ||
Db       731 GCGGCGCCGCGCCCCGCCCGCG---CCGNCGCCGCGCCGCCGCGGGGCCCCGCCCGCGGC 675

Qy        659 GGAGATTCGCCTCCCGGAGCTTCACCACG--TGGACTTCTCCACCTTTGCCCTCCATCCC 716
              || |   ||| || || ||  | ||  |  | |  |    |  |  | || |
Db       674 GGCGCGGGGCCCCCGGGGGCCGCGCCGGGGCGCGGCGCCGGGCCGGCGGGGCGCCCGCGG 615

Qy        717 GCCCTGCTCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGA 776
              |  |    |  |||  ||   |     || ||||| || |  ||  | |  |||||
Db       614 GGGCGCGCCCCCGCGCCCGCCCGGGCGNCGGGCGGGCGGCGGCGCGCCGCGCGGGCCCCC 555

Qy        777 GCCCCCTATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGG 836
              | || |   |  ||   | |||    || ||| | | |    || | |   |  ||| |
Db       554 GGCCGCGCCCCCCCCCCCCCCCCGCGCGGGGGCCCCCCNNCCCNCNCCGGGGGGGCCCCC 495

Qy        837 CTCTTCGCCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCAC 896
              | |   ||   || | ||  |  || | || ||   | |   |  |   || | | |
Db       494 CCCCCNNCCCNCNCCNCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNNCCCCCCCCCCCC 435

Qy        897 GTCACGCTGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGC 956
              | | |   ||  ||| ||     | | |  |       | |  ||  | |
Db       434 CCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCNCCNCCCCCCCCCCCCCCCCCCNCCC 375

Qy        957 CTCGCGACGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTC 998
              | | |  |  |  |  | | ||| |||||  |||  |  | |
```

```
Db        374 CCCCCCCCNNCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCNC 333


RESULT 15
AG435170/c
LOCUS       AG435170               1822 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-313C14.TJ, genomic survey
            sequence.
ACCESSION   AG435170
VERSION     AG435170.1  GI:48078233
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1822)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1822
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-313C14.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             9.7%;  Score 97.2;  DB 10;  Length 1822;
  Best Local Similarity   44.2%;  Pred. No. 7.7e-09;
  Matches  440;  Conservative    0;  Mismatches  543;  Indels   12;  Gaps    2;

Qy         11 GCCGCTCGACGCGCGCCGGGCGCGGCGGCGCGTCCCGCTGCCCACGTACCCCTTCGCCCG 70
              ||  | |  |   | |||| |||| |  |||  ||  |   || |   ||||  ||| |
Db       1723 GCGCCCCCCCCGCCCCCGGCCGCGCCCCCGCCCCCGCCGCGCCGCNCCCCCCGCCGCGCN 1664

Qy         71 CGAGCGTTACTGGTTCTCCAGGAGCGGAGACGCCTTCACCCTCGGCCAGGCGGGAGAGAG 130
              ||  |    |     | |  |  | | |   |     ||  ||  |  ||||  | |
Db       1663 CGCNCCGGCCGCCCGCGCGGCGCCCCGCGCGCCGCGCCCGCCGCGCGCNCGCGGCCGCGCC 1604

Qy        131 GCGCTTGCACCCGCTCGTGCAGGCGAACACCTCCACGTTCCACGCGCACACGTACTCCAG 190
              |||    | | ||  |  || | |       | || |   || ||||| | ||  | |
Db       1603 GCGNCCCCCCGCGGCCCCGCCGCCCCCGCGCCCCNCCGCCCCCGCGCCCGCGGGCGCGCC 1544

Qy        191 CCGGCTCCGGGGGCGACGCGTTCTACCTCGCCGATCACCTCGTGCACGGCCAGAAGCTCCT 250
              ||| |  ||   | |||   |  |   |     | || |   |    ||       |
Db       1543 CCGCCCGCGCNCCCCCGCCGCCCCGCCGCCGNCCCNCCGCCCCCCGCCCCCCGCCGCCGC 1484

Qy        251 CCCCGCGGCGGCGTTCCTGGAGATGGCCCGCGCCGCCGGGGAGATGGCGTCCGGGCGGCC 310
              |||||| ||| |   ||      |||| ||||     | |   ||  ||  |  ||
Db       1483 CCCCGCCGCGCCCCCCCCCGCGCCGCCCCCGCCC-----GNGCCCCCGCGCGCCCCCCC 1429

Qy        311 GGTCCGCGACATCCTCGACGTCGTCTGGACCGCGCCCGTCGTCGTGGGCGCCGAGCCGCG 370
              |  ||  |   |  ||    |  || |    |||| |||| |   |  |  |||| |||  |
Db       1428 GCCCCCCCCCCCCCCGCGCCCGCCCNCCCCGCCCCCGCCCGCCCGCCCGCCCCGCCCNG 1369

Qy        371 CGAGATCGAGATAACGCTCCGGCCGGCCGCCGGCGCCATCGACTTCGCCGTGTCCTCCGC 430
              |  |     |   ||  |  ||  ||     | ||  |  |     |   || ||||
Db       1368 CCCGCCGCCGCCCCGGCCGCCCCCCCCCCGGCCCCCCCGCCGCCCGCGCCGCGCCCCCGC 1309

Qy        431 CGCCGAGCGCGCGGTGATCTCCCACGCGCAGGGGCGGATGCGCCTCGACGAGGGGGATCC 490
              |  |  |||| |      |  || | |  | ||     | || |     ||     ||
Db       1308 CCGCCCGCGCCCCCCCCCCCGCCNCCCCCCCGCGCNCCGCCCCCGCCGGCCGGCCCCCCC 1249

Qy        491 CGCCGAAGAGGCGGCGCCGCCCCTCCCGCTCGATGACATCCTCTCAC-------GTTGCT 543
              ||||      | ||||||| ||||  |  | |   | |  || | |           |   |
Db       1248 CGCCCCGCCGCCGGCGCCCCCCCGGCGCCCCCCGGNCCCCCGCGCGNCCCCCCGGGCCCG 1189

Qy        544 CGAGGGTCACCGGCGGAGACGCGTGCTACCGCCGCCTCCAGCAGCTCGGGCTGCACCACG 603
              ||   | | ||| |     |||   |  |||||||| ||  | ||  | | | || || |
Db       1188 CGCCCGCCCCCGCCCCCCGCGCCGCCGCCCGCCGCCCCCCCCCGCGCCGCCGGCGCCGCC 1129

Qy        604 GCGGCAGCATGCGCGCGCTCCACGAGCTGCGCCGAGGCGAGGGCGAGGCCATCGCGGAGA 663
              |  | ||  | ||| | || |   |  |  |    ||  | |   | ||  | | | |
Db       1128 CCCCCCGCCGCCCCGCCCCCCCCCCCCGCCCGCCGCCCGCCGCCCCCGNCCCCCCCCGCGC 1069

Qy        664 TTCGCCTCCCGGAGCTTCACCACGTGGACTTCTCCACCTTTGCCCTCCATCCCGCCCTGC 723
              | |  ||||   |  | ||| |          | || |      || ||  ||| ||| |
Db       1068 CCCCCGCCCCGCCCCCCCCCCCCCGCCCGCCCCCCCCGCCGCGCCCCCGCCCCCCCCGG 1009

Qy        724 TCGACGCTGCCCTGCAATGCACGCTCGGGCTGCTGGACGATGAGGCGGCCCGAGCCCCCT 783
              |  | | |||   |   ||  | |  ||          |    ||   ||  |||||
Db       1008 CCCGCCCGGCCGCCCCGCGCCGCCCCCCGCCCGCCCCCCCCCCCGCCCNCCCCCCCCCCC 949

Qy        784 ATCTTCCTTTCGCCGTCGGCCGGGTCACGCTGCTCCGCCCGCTGCCGGCGCGGCTCTTCG 843
                 ||   | ||  |  ||    |  | |  || ||| | ||| || || | |  ||
Db       948 GCNCCCCGCGCCCCCCCCCCCCCCCCCCGNCCGNGCCCCCCCCCGCCCGCCCGCCCCCCG 889

Qy        844 CCTATGCCACGCCGTCGTCCGCGCCGCCGGGCACGAACGCCAGGGCCTCTCACGTCACGC 903
              ||    | | ||  |  |  ||  | || | | |   ||||  | || | | || | | |
Db       888 CCCCCCGCCCCCCCGCCCCCCGCCGCCCCGCCGCCCGCGCCGCGCCCCCCCGCGCCCCCC 829

Qy        904 TGGCCGATCCCGCCGGCCGGGTGCTCCTCGAGATGCGTGATTTCACCGTCCGCCTCGCGA 963
              ||  | |  |  | |    || |      |       | | | ||||| ||||
Db       828 CCCCCCCCCGCCGCCCCGCGCCCGCCCCCNCCCNCCCCCCCCNCCCGGGCCGCCGCGCGG 769

Qy        964 CGGCGGACGTCGCGCCCACCCCCGCCCAGCGGCTC 998
              |  |   || | | ||| || ||   | ||  | |
Db       768 CCCCCCCCGGCCCCCCCCCCGCCCGNCCGCCCCGC 734


Search completed: April 28, 2006, 12:17:47
Job time : 4958.14 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 5546.62 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 cgccacgagcgcgaaggcgg.....accgccggccgcctcgaagc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1001	100.0	84868	1	DQ013294	DQ013294 Polyangiu
	2	137.6	13.7	9867	6	AR619613	AR619613 Sequence
c	3	135.8	13.6	2973	6	AR628165	AR628165 Sequence
c	4	113.8	11.4	6979	1	AB045630	AB045630 Pseudomon
c	5	113.8	11.4	10929	1	AE004886	AE004886 Pseudomon
	6	109.6	10.9	159997	14	CR352334	CR352334 Danio rer
c	7	107.4	10.7	1677	6	AX078627	AX078627 Sequence
	8	107.2	10.7	1393	10	PM11H12G	AL684264 Penicilli
c	9	107.2	10.7	4957	1	AY929300	AY929300 Pseudomon
c	10	107.2	10.7	300933	1	AE016791	AE016791 Pseudomon
	11	105	10.5	956	10	PM2D12B	AL684743 Penicilli
c	12	104.4	10.4	207592	14	CR392332	CR392332 Danio rer
	13	102.8	10.3	110000	1	AE016853_10	Continuation (11 o
c	14	101.4	10.1	1065	10	PM2B12B	AL684695 Penicilli
c	15	99.4	9.9	1393	10	PM11H12G	AL684264 Penicilli
c	16	99.2	9.9	110000	1	CP000076_60	Continuation (61 o
	17	98.2	9.8	1094	10	PM7G11B	AL685196 Penicilli
c	18	98	9.8	186620	14	CR405691	CR405691 Danio rer
c	19	96.4	9.6	159997	14	CR352334	CR352334 Danio rer
c	20	96.4	9.6	227214	1	BX572608	BX572608 Rhodopseu
	21	94.8	9.5	110000	1	CP000075_09	Continuation (10 o
	22	94.4	9.4	1007	10	PM3H11G	AL685749 Penicilli
c	23	93.8	9.4	65351	14	AC139773	AC139773 Homo sapi
	24	93.6	9.4	1279	10	PM2H12G	AL684840 Penicilli
	25	93.4	9.3	110000	1	CR555306_00	CR555306 Azoarcus
	26	93.4	9.3	110000	1	CR555306_01	Continuation (2 of
c	27	93.2	9.3	1094	10	PM7G11B	AL685196 Penicilli
	28	93.2	9.3	136753	14	AC149968	AC149968 Strongylo
c	29	93.2	9.3	136753	14	AC149968	AC149968 Strongylo
	30	93.2	9.3	156052	14	CR749776	CR749776 Danio rer
	31	93	9.3	82746	1	AF453501	AF453501 Actinosyn
c	32	91.6	9.2	237821	14	CR855879	CR855879 Danio rer
	33	91.2	9.1	1065	10	PM2B12B	AL684695 Penicilli
c	34	90.8	9.1	1377	10	PM12C6G	AL684346 Penicilli
	35	90.4	9.0	1392	10	PM12B10G	AL684308 Penicilli
	36	90	9.0	110000	1	CP000058_10	Continuation (11 o
	37	89.6	9.0	186620	14	CR405691	CR405691 Danio rer
	38	89.6	9.0	290452	14	AC079167	AC079167 Mus muscu
c	39	89.4	8.9	1143	10	PM12A11G	AL684286 Penicilli
c	40	89.4	8.9	110000	1	CP000031_10	Continuation (11 o
	41	89.4	8.9	150715	13	AY714813	AY714813 Cercopith
c	42	89.4	8.9	150715	13	AY714813	AY714813 Cercopith
	43	89	8.9	1377	10	PM12C6G	AL684346 Penicilli
	44	89	8.9	202971	14	CR759948	CR759948 Danio rer
	45	89	8.9	272101	1	AE017302	AE017302 Thermus t

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRAREAEREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPFFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

```
                GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
     CDS        complement(1992. .3440)
                /note="ORFu2; similar to Pseudomonas putida KT2440 PP4696;
                putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32957.1"
                /db_xref="GI:63099934"
                /translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
                APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
                QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
                KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
                EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
                VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
                ELTETEIARRLGISRKALWERRQRLGIPRPRS"
     CDS        3587. .5548
                /note="ORFu3; ORFu3p; similar to Leptospira interrogans
                AAN50474"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32958.1"
                /db_xref="GI:63099935"
                /translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
                VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
                LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
                YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
                SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
                IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
                VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
                AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
                GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
                AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
                SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
                VVAGFGATGPLRVMETLPEQYLFLTRVRP"
     CDS        complement(5702. .7231)
                /note="ORFu4; ORFu4p; similar to Burkholderia fungorum
                hypothetical protein in ZP_00033665"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32959.1"
                /db_xref="GI:63099936"
                /translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
                AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
                PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
                VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
                MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
                QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPFSANQRPASTLAQWRVGLEHP
                GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
                GRVITVHFHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
                ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
                P"
     CDS        complement(7561. .8931)
                /note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440
                PP4696; putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32960.1"
                /db_xref="GI:63099937"
                /translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
                APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
                AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
                VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
                VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
                PPPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
                IPRPRS"
     CDS        complement(8928. .11939)
                /note="ORF1; ORF1p; similar to Pseudomonas putida KT2440
                HisK (PP4695) in GenBank Accession Number AAN70268"
                /codon_start=1
                /transl_table=11
                /product="putative sensory box histidine kinase"
                /protein_id="AAY32961.1"
                /db_xref="GI:63099938"
                /translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
                YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
                AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
                FTAFFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
                HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
                FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
                VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
                LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
                IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPAREEREAAR
                LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
                LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
                RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
                LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
                GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
                DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
                DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
                IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
                LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
     CDS        complement(12405. .13379)
                /note="ORF2; ORF2p; similar to family of known or putative
                phosphotransferases, including macrolide
                2-phosphotransferases; similar to Bacillus subtilis YcbJ,
                Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273
                MphB and Staphylococcus aureus plasmid pSR1 MphBM"
                /codon_start=1
                /transl_table=11
                /product="putative macrolide kinase"
                /protein_id="AAY32962.1"
                /db_xref="GI:63099939"
                /translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
                DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
                VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
                IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
                DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
                AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
     CDS        13368. .14663
                /note="ORF3; ORF3p"
                /codon_start=1
                /transl_table=11
                /product="putative Ser/Thr protein kinase"
                /protein_id="AAY32963.1"
                /db_xref="GI:63099940"
                /translation="MFGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
                ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
                RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
                GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
                AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
                DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
                TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
                LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
     gene       15731. .33766
                /gene="dszA"
     CDS        15731. .33766
                /gene="dszA"
                /note="PKS (polyketide synthase) subunit modules 1-4a"
                /codon_start=1
                /transl_table=11
                /product="DszA"
                /protein_id="AAY32964.1"
                /db_xref="GI:63099941"
                /translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
                PRAHYSPDFNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
                IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT
```

```
                     LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                     LSPDGLCKTFDADANGYVPGDGVGVLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                     ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                     RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                     VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                     RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                     DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                     RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                     RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                     ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                     RAQFTFKKFLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                     LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                     PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                     PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                     EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE

  Query Match          100.0%;  Score 1001;  DB 1;  Length 84868;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-81;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11574 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 11633

Qy         61 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11634 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 11693

Qy        121 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11694 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 11753

Qy        181 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11754 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 11813

Qy        241 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11814 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 11873

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11874 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 11933

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11934 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 11993

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11994 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 12053

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12054 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 12113

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12114 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 12173

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12174 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 12233

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12234 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 12293

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12294 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 12353

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12354 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 12413

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12414 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 12473

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12474 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 12533

Qy        961 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAGC 1001
              |||||||||||||||||||||||||||||||||||||||||
Db      12534 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAGC 12574


RESULT 2
AR619613
LOCUS       AR619613                9867 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 1008 from patent US 6833447.
ACCESSION   AR619613
VERSION     AR619613.1  GI:59724019
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 9867)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 1008 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .9867
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match          13.7%;  Score 137.6;  DB 6;  Length 9867;
  Best Local Similarity  54.5%;  Pred. No. 0.00036;
  Matches  300;  Conservative    0;  Mismatches  244;  Indels    6;  Gaps    1;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              || | |||  |||    |  ||   ||  |||| ||   | | | || ||||||| || |
Db       3497 CCGCCAGCAGGAACAGTGTCACCGCCGTACCGGTGGATTGCCCCGGGTAGCGGAAGGCCA 3556

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              ||| ||| |  |||||||  |||   ||  |||||| |  | ||||||||    |||  |
Db       3557 GCACGTCCGCTAGCGACGTGAGCTGCAACTCGCGGGACAGGCGCAGCAGCGGGCGCCACA 3616

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | |  ||  |  || |||||| || ||   ||| |||| |   | ||| ||  ||  |
Db       3617 GCACGGGTATCAGCAGGCACGCCAGGGTGACGCCCAGGTAGATGCCCAGGTAGCGGAAGC 3676

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              | || ||   |||| |||| |||| |||| |||| | |||||| | |||||| |||||  |
Db       3677 CATGCCGGGCCGCGTAGCCCACGCTGCCGAAGTAGGACCACGATGTGGCGTACACGCCCA 3736

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              | |  || |||    ||||||||   | | | ||         |||    | | |||| ||
Db       3737 GGGCCAGCGCGTAGACGAGCGGATGCTGGGTGATGCGCGCGGGGATGGCGCCTCGCTCCG 3796

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              | |||   || | |  |  ||||||    ||  | || || |  |||   | |||   |
Db       3797 CCGCGTAGGCCACCAGGAAGAGCACCAGCAGGTAGGCGACGGACGCGGCGACGAGCGAAC 3856

Qy        363 CCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCG 422
              | |  |||| |      ||       |     | |||   ||  | |  | | |   | |
Db       3857 CGAGCTCGAGCGTCATGG------GCCACGACTCCGGTGCGCCGCCCAGCCGCCCAGGAG 3910

Qy        423 CGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGC 482
              |  ||  |      | ||||   ||  |  ||  |  ||||    | |||  |  |
Db       3911 GATGACGCCCAGCCACACCGCGAAGGGCAGGTACCCCGGCGCGCCCTCCGCCAGCCACAG 3970
```

```
Qy        483 GAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAG 542
              |  |   | | |  | ||  | | || |||     ||| || || |     | |||
Db       3971 TCGGCGCAGGGGAGAGCCGAAGAGGAGCACCGCGACGGCGAAGACGAGCAGCGACGGCAT 4030

Qy        543 CACGAGCACG 552
              |||| |  ||
Db       4031 CACGGGGGCG 4040


RESULT 3
AR628165/c
LOCUS       AR628165               2973 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 9566 from patent US 6833447.
ACCESSION   AR628165
VERSION     AR628165.1  GI:59736977
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2973)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 9566 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .2973
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             13.6%;  Score 135.8;  DB 6;  Length 2973;
  Best Local Similarity   60.4%;  Pred. No. 0.00079;
  Matches  224;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              || | |||  |||    |  ||   ||  |||| ||   | | | || |||||| || |
Db        376 CCGCCAGCAGGAACAGTGTCACCGCCGTACCGGTGGATTGCCCCGGGTAGCGGAAGGCCA 317

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              ||| ||| |  |||||| ||  ||  |||||| |  | |||||||    |   ||| |
Db        316 GCACGTCCGCTAGCGACGTGAGCTGCAACTCGCGGGACAGGCGCAGCAGCGGGCGCCACA 257

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | |  ||  |   || |||||| || ||   ||| ||||| |   | ||| ||  ||  |
Db        256 GCACGGGTATCAGCAGGCACGCCAGGGTGACGCCCAGGTAGATGCCCAGGTAGCGGAAGC 197

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              | || ||   |||| |||| |||| |||| | |||||| | |||||| ||||| |
Db        196 CATGCCGGGCCGCGTAGCCCACGCTGCCGAAGTAGGACCACGATGTGGCGTACACGCCCA 137

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              | |  || |||   |||||||   | | | ||         |||     | | |||| ||
Db        136 GGGCCAGCGCGTAGACGAGCGGATGCTGGGTGATGCGCGCGGGGATGGCGCCTCGCTCCG 77

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              | |||   || | |  | ||||||    ||  | || || |  |||   | |||    |
Db         76 CCGCGTAGGCCACCAGGAAGAGCACCAGCAGGTAGGCGACGGACGCGGCGACGAGCGAAC 17

Qy        363 CCACGTCGATC 373
              | |  |||| |
Db         16 CGAGCTCGAGC 6


RESULT 4
AB045630/c
LOCUS       AB045630               6979 bp    DNA     linear   BCT 19-JUN-2001
DEFINITION  Pseudomonas aeruginosa genes for glutamate-tRNA synthetase, CbrA,
            CbrB, complete cds.
ACCESSION   AB045630
VERSION     AB045630.1  GI:12082248
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (sites)
  AUTHORS   Nishijyo,T., Haas,D. and Itoh,Y.
  TITLE     The CbrA-CbrB two-component regulatory system controls the
            utilization of multiple carbon and nitrogen sources in Pseudomonas
            aeruginosa
  JOURNAL   Mol. Microbiol. 40 (4), 917-931 (2001)
   PUBMED   11401699
REFERENCE   2  (bases 1 to 6979)
  AUTHORS   Itoh,Y. and Nishijyo,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Yoshifumi Itoh, National Food Research
            Institute, Applied Microbiology; Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8642, Japan (E-mail:yosifumi@nfri.affrc.go.jp,
            Tel:81-298-38-8075, Fax:81-298-38-7996)
FEATURES             Location/Qualifiers
     source          1. .6979
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            284. .1165
                     /gene="gltX"
     CDS             284. .1165
                     /gene="gltX"
                     /note="putative"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="glutamate-tRNA synthetase"
                     /protein_id="BAB20865.1"
                     /db_xref="GI:12082249"
                     /translation="MTSSYVGRFAPTPSGYLHFGSLVAAVASYLDARAVGGRWLVRME
                     DLDPPREVPGAQQAILETLERYGFEWDGAVERQSERFPAYAAVVEQLLRSGLAYACTC
                     SRKQLEGFAGIYPGFCRDAGHAREDAAIRLRVPELEYRFVDRVQGEVRQHLGREVGDF
                     VIQRRDGLYAYQLAVVLDDAWQGITDIVRGADLLDSTPRQLYLQELLGLSQPRYLHVP
                     LIVQPDGHKLGKSYRSPPLPAEQAAAPLTRALRALGQRPPAELAQASASEALAWGVAH
                     WDATRIPRCATLPEERL"
     gene            1395. .4346
                     /gene="cbrA"
     CDS             1395. .4346
                     /gene="cbrA"
                     /function="signal sensor and protein histidine kinase"
                     /note="KinA-NtrB family
                     histidine at 766, phosphrylation site
                     histidine kinase domain,491 to 983 amino acids
                     sensor membrane domain, 1 to 490 amino acids"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CbrA"
                     /protein_id="BAB20866.1"
                     /db_xref="GI:12082250"
                     /translation="MLTSFSLTQLILISVTYLSTLFGIAWITERGYVPRRLVRHPLVY
                     TLSLGVYASAWAFYGTVGLAYQYGYGFLAIYLGISGAFLLAPVLLYPILRVTRAYQLS
                     SLADLFAFRFRSTWAGALTTLFMLIGVLPMLALQIQAVTDSISILTRDPAQERVAFVF
                     CTLITLFAILFGARHIATRERHEGLVFAIAFESLVKLVMLGSIGLYALYGVFGGPEGL
                     EVWLLQNQTALTTLHTPLAEGPWRTLLLVFFAAAIVMPHMFHMIFTENLNPRALISAS
                     WGLPLFLLLMSLAVPPILWAGLRLGASTTPEYYTIGLGLAVDSPALALAAFIGGISAA
                     SGLTIVMTLALSGMVLNHLVLPLYQPPAQGNIYRWLKWTRRLLIVAIIMASYAFYLLL
                     GAEQDLSNLGIVSFVATLQFLPGALSVLYWPTANRRGFISGLIAGMLVWAITMLLPLM
                     GNVQGLYLPLFDVIYVLDDSNWHLAALSSLAANVLVFTLVSLFTEASDEEKGAAEACA
                     VDNVRRPQRRELFAGSPQEFASQLAKPLGAKTAQKEVEQALRDLHLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDIVETFLPYKTSEESYVTEDIHFIESRLEDYHSRLTGLA
                     AELDTLRRYHRQTLQDLPMGVCSLAKDQEVLMWNRAIEELTGVGAQKVVGSRLSALPE
                     PWKGLLERFIDAPDEHLHKQRLVYDGHTRWLNLHKAAIEEPLAPGNSGLVLLVEDLTE
                     TQLLEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGELTEISEQ
                     ILDQTKRVSRIVQSLMSFAHSGSHLQALEPVCLSEVAQEAIGLLSLNRRSVEVEFFNL
                     CDPAHWVEGDSQRLAQVLINLLSNARDASPPGGAIRVRSEASEHTVDLVVEDEGSGIP
                     KAIMDQLFEPFFTTKDPGKGTGLGLALVYSIVEEHYGQITIESPTDHQREGGTRFRVT
                     LPRHPGPTAEL"
     gene            4373. .5809
```

```
                     /gene="cbrB"
     CDS             4373. .5809
                     /gene="cbrB"
                     /function="response regulator"
                     /note="Aspertate at 52 amino acids, phosphorylation site
                     NtrC family"
                     /codon_start=1
                     /transl_table=11
                     /product="CbrB"
                     /protein_id="BAB20867.1"
                     /db_xref="GI:12082251"
                     /translation="MAHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERYTIP
                     SFDMVVSDLRLPGAPGTELIKLAEGIPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILKDRQENRSAAPVSANGGKAGGERGASPAVADGEIGIIGSCAPMQELY
                     SKIRKVAPTDSTVLIQGESGTGKELVARALHNLSKRAKAPLISVNCAAIPETLIESEL
                     FGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRVGSV
                     QSQKVDVRLIAATHRDLKTLAKTGQFREDLYYRLHVISLKLPALRERGNDVMEIARAF
                     LARQCTRMGRAPLSFAHDAEQAIRHYPWPGNVRELENAIERAVILSESQEIHADLLGI
                     DIELDDLEDDFAADLGLGPAGAAASNHEPTEDLSLEDYFQHFVLEHQDHMTETELARK
                     LGISRKCLWERRQRLGIPRRKSGASADS"
ORIGIN

  Query Match           11.4%;  Score 113.8;  DB 1;  Length 6979;
  Best Local Similarity  57.0%;  Pred. No. 0.052;
  Matches  208;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  |||    | |   || || |||||   ||  ||  |||| ||||| ||||
Db       1776 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 1717

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
               |||||||  ||||| || ||  | |   |||| || |  ||||| |   |       |
Db       1716 AGAGGTCGGCCAGCGAGGACAGCTGGTAGGCGCGTGTCACCCGCAGGATCGGATAGAGCA 1657

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| | |  || | | ||    |     || ||||| |  |||||||  ||   |
Db       1656 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 1597

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   ||||||| ||| ||| |    ||||||||||| |
Db       1596 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 1537

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |   || ||||||   ||  | |||  ||||||       | ||||||  |
Db       1536 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 1477

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              |     |||||  ||  |||   |  |||  | |  ||  ||   |  |  ||   |
Db       1476 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGTGACGCTGATCAGGATCAGCTGGG 1417

Qy        363 CCACG 367
                 || |
Db       1416 TCAGG 1412


RESULT 5
AE004886/c
LOCUS       AE004886               10929 bp    DNA     linear   BCT 19-FEB-2003
DEFINITION  Pseudomonas aeruginosa PAO1, section 447 of 529 of the complete
            genome.
ACCESSION   AE004886 AE004091
VERSION     AE004886.1  GI:9950982
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa PAO1
  ORGANISM  Pseudomonas aeruginosa PAO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 10929)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
   PUBMED   10984043
REFERENCE   2  (bases 1 to 10929)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 10929)
  AUTHORS   .
  CONSRTM   Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2003) Department of Molecular Biology and
            Biochemistry, Simon Fraser University, 8888 University Dr.,
            Burnaby, British Columbia V5A 1S6, Canada
COMMENT     ----------------------------------------------------------------
            This represents the February 3, 2003 version of the continually
            updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,
            from PseudoCAP (see http://www.pseudomonas.com for latest updates
            and links to alternate annotations). PseudoCAP is coordinated by
            Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
            E.W. Hancock (University of British Columbia, Canada). We welcome
            submission through www.pseudomonas.com of any proposed changes.

            'Protein name confidence' is used to rate our confidence of the
            accuracy of the protein name.
            Class 1: Function experimentally demonstrated in P. aeruginosa.
            Class 2: Function of highly similar gene experimentally
                  demonstrated in another organism (and gene context consistent
                  in terms of pathways its involved in, if known).
            Class 3: Function proposed based on presence of conserved amino
                  acid motif, structural feature or limited sequence similarity
                  to an experimentally studied gene.
            Class 4: Homologs of previously reported genes of unknown function,
                  or no similarity to any previously reported sequences.
            ----------------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .10929
                     /organism="Pseudomonas aeruginosa PAO1"
                     /mol_type="genomic DNA"
                     /strain="PAO1"
                     /db_xref="taxon:208964"
     gene            127. .573
                     /gene="dksA"
                     /locus_tag="PA4723"
     CDS             127. .573
                     /gene="dksA"
                     /locus_tag="PA4723"
                     /note="Protein name confidence: Class 1 (function
                     experimentally demonstrated in P. aeruginosa)"
                     /codon_start=1
                     /transl_table=11
                     /product="suppressor protein DksA"
                     /protein_id="AAG08109.1"
                     /db_xref="GI:9950983"
                     /translation="MSTKAKQQSSQQMTRGFEPYQETKGEEYMSERMRAHFTAILNKW
                     KQELMEEVDRTVHHMQDEAANFPDPADRASQEEEFSLELRARDRERKLIKKIDETLQL
                     IEDEEYGWCDSCGVEIGIRRLEARPTATLCIDCKTLAEIREKQLGS"
     gene            641. .1522
                     /locus_tag="PA4724"
                     /note="synonym: yadB"
     CDS             641. .1522
                     /locus_tag="PA4724"
                     /note="Protein name confidence: Class 3 (function proposed
```

based on presence of conserved amino acid motif,
structural feature or limited sequence similarity to an
experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable aminoacyl-transfer RNA synthetase
(class I)"
/protein_id="AAG08110.1"
/db_xref="GI:9950984"
/translation="MTSSYVGRFAPTPSGYLHFGSLVAAVASYLDARAVGGRWLVRME
DLDPPREVPGAQQAILETLERYGFEWDGAVERQSERFPAYAAVVEQLLRSGLAYACTC
SRKQLEGFAGIYPGFCRDAGHAREDAAIRLRVPELEYRFVDRVQGEVRQHLGREVGDF
VIQRRDGLYAYQLAVVLDDAWQGITDIVRGADLLDSTPRQLYLQELLGLSQPRYLHVP
LIVQPDGHKLGKSYRSPPLPAEQAAAPLTRALRALGQRPPAELAQASASEALAWGVAH
WDATRIPRCATLPEERL"

gene 1752. .4703
/gene="cbrA"
/locus_tag="PA4725"

CDS 1752. .4703
/gene="cbrA"
/locus_tag="PA4725"
/note="Protein name confidence: Class 1 (function
experimentally demonstrated in P. aeruginosa)"
/codon_start=1
/transl_table=11
/product="two-component sensor CbrA"
/protein_id="AAG08111.1"
/db_xref="GI:9950985"
/translation="MLTSFSLTQLILISVTYLSTLFGIAWITERGYVPRRLVRHPLVY
TLSLGVYASAWAFYGTVGLAYQYGYGFLAIYLGISGAFLLAPVLLYPILRVTRAYQLS
SLADLFAFRFRSTWAGALTTLFMLIGVLPMLALQIQAVTDSISILTRDPAQERVAFVF
CTLITLFAILFGARHIATRERHEGLVFAIAFESLVKLVMLGSIGLYALYGVFGGPEGL
EVWLLQNQTALTTLHTPLAEGPWRTLLLVFFAAAIVMPHMFHMIFTENLNPRALISAS
WGLPLFLLLMSLAVPPILWAGLRLGASTTPEYYTIGLGLAVDSPALALAAFIGGISAA
SGLTIVMTLALSGMVLNHLVLPLYQPPAQGNIYRWLKWTRRLLIVAIIMASYAFYLLL
GAEQDLSNLGIVSFVATLQFLPGALSVLYWPTANRRGFISGLIAGMLVWAITMLLPLM
GNVQGLYLPLFDVIYVLDDSNWHLAALSSLAANVLVFTLVSLFTEASDEEKGAAEACA
VDNVRRPQRRELFAGSPQEFASQLAKPLGAKTAQKEVEQALRDLHLPFDERRPYALRR
LRDRIEANLSGLMGPSVAQDIVETFLPYKTSEESYVTEDIHFIESRLEDYHSRLTGLA
AELDTLRRYHRQTLQDLPMGVCSLAKDQEVLMWNRAIEELTGVGAQKVVGSRLSALPE
PWKGLLERFIDAPDEHLHKQRLVYDGHTRWLNLHKAAIEEPLAPGNSGLVLLVEDLTE
TQLLEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGELTEISEQ
ILDQTKRVSRIVQSLMSFAHSGSHLQALEPVCLSEVAQEAIGLLSLNRRSVEVEFFNL
CDPAHWVEGDSQRLAQVLINLLSNARDASPPGGAIRVRSEASEHTVDLVVEDEGSGIP
KAIMDQLFEPFFTTKDPGKGTGLGLALVYSIVEEHYGQITIESPTDHQREGGTRFRVT
LPRHPGPTAEL"

gene 4730. .6166
/gene="cbrB"
/locus_tag="PA4726"

CDS 4730. .6166
/gene="cbrB"
/locus_tag="PA4726"
/note="Protein name confidence: Class 1 (function
experimentally demonstrated in P. aeruginosa)"
/codon_start=1
/transl_table=11
/product="two-component response regulator CbrB"
/protein_id="AAG08112.1"
/db_xref="GI:9950986"
/translation="MAHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERYTIP
SFDMVVSDLRLPGAPGTELIKLAEGIPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
DEMLQAVARILKDRQENRSAAPVSANGGKAGGERGASPAVADGEIGIIGSCAPMQELY
SKIRKVAPTDSTVLIQGESGTGKELVARALHNLSKRAKAPLISVNCAAIPETLIESEL
FGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRVGSV
QSQKVDVRLIAATHRDLKTLAKTGQFREDLYYRLHVISLKLPALRERGNDVMEIARAF
LARQCTRMGRAPLSFAHDAEQAIRHYPWPGNVRELENAIERAVILSESQEIHADLLGI
DIELDDLEDDFAADLGLGPAGAAASNHEPTEDLSLEDYFQHFVLEHQDHMTETELARK
LGISRKCLWERRQRLGIPRRKSGASADS"

gene 7067. .8470
/gene="pcnB"
/locus_tag="PA4727"

CDS 7067. .8470
/gene="pcnB"
/locus_tag="PA4727"
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein); Alt protein name:
polynucleotide adenylyltransferase, plasmid copy number
protein"
/codon_start=1
/transl_table=11
/product="poly(A) polymerase"
/protein_id="AAG08113.1"
/db_xref="GI:9950987"
/translation="MLKKLIQTIRSPLRRPRAVRTTPEVIGNNQHSLRRDQFSRNAVK
VVEILQRAGYQAYVVGGCVRDQMLGIAPKDFDVATSATPEQVRAEFRNARIIGRRFKL
VHVHFGREIIEVATFRAHHPEGQDDGDSRASSNESGRILRDNVYGSLEDDAQRRDFTI
NALYFDVTSERLLDYANGVHDIRNRLIRLIGDPEQRYLEDPVRMLRAVRFAAKLDFEI
EKHSAAPIRRLAPLLREIPSARLFDEVLKLFLAGRAERTFELLVEYELFAPLFPASAK
ALQANPDYTGKLIRQALANTDARIRQGKPVTPAFLFAALLWPALPARVAQLQEKGMPA
IPAMQEAAHELISEQCQRIAIPKRFTLPIREIWDMQERLPRRQGKRADLLLENPRFRA
GYDFLLLRESAGEETEGLGQWWTDYQEVSDSERRNMIRDLVSQEDGSAPRKRRRGGNS
GRRRRGPRKEGSGGSGE"

gene 8467. .8955
/gene="folK"
/locus_tag="PA4728"

CDS 8467. .8955
/gene="folK"
/locus_tag="PA4728"
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein); Alt protein name:
6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase,
7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase"
/codon_start=1
/transl_table=11
/product="2-amino-4-hydroxy-6-
hydroxymethyldihydropteridine pyrophosphokinase"
/protein_id="AAG08114.1"
/db_xref="GI:9950988"
/translation="MIQRVYVALGSNLAEPREQIQAALDAFERLPETRLVAVSPLYIS
DPLGPADQPRFVNGVAALDTNLAPLDLLDALQAIELEQGRVRDLRWGPRTLDLDILLF
GEQLLDLPRLKVPHYHMQARAFVLYPLADLAPDLRLPDGRHLPELLAACPFEGIERLP

Query Match 11.4%; Score 113.8; DB 1; Length 10929;
Best Local Similarity 57.0%; Pred. No. 0.045;
Matches 208; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

```
Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  |||     | |    || || |||||    ||  ||  |||| ||||| ||||
Db       2133 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 2074

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
                |||||||  |||||  || ||  | |    |||| || |  |||||  |    |       |
Db       2073 AGAGGTCGGCCAGCGAGGACAGCTGGTAGGCGCGTGTCACCCGCAGGATCGGATAGAGCA 2014

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| | |  || | | ||     |     || |||||  |  ||||||||  ||   |
Db       2013 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 1954

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   |||||||  ||| ||| |    ||||||||||| |
Db       1953 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 1894

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |   || ||||||   ||  | |||  ||||||       | ||||||  |
Db       1893 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 1834

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              |      |||||  ||  |||   |  |||  | |  ||  ||   |  |  ||   |
Db       1833 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGTGACGCTGATCAGGATCAGCTGGG 1774

Qy        363 CCACG 367
                |||
Db       1773 TCAGG 1769
```

RESULT 6

```
CR352334
LOCUS       CR352334              159997 bp    DNA     linear   HTG 09-FEB-2005
DEFINITION  Danio rerio clone DKEY-221H20, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR352334
VERSION     CR352334.6  GI:58865138
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 159997)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 10, 2005 this sequence version replaced gi:46559209.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK221H20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 156087 bases at least Q40
            Consensus quality: 156644 bases at least Q30
            Consensus quality: 157332 bases at least Q20
            Insert size: 159297; sum-of-contigs
            Insert size: 160605; 1.3% error; agarose-fp
            Quality coverage: 12.78x in Q20 bases; sum-of-contigs Quality
            coverage: 12.68x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    38359: contig of 38359 bp in length
            *    38360    38459: gap of 100 bp
            *    38460    42262: contig of 3803 bp in length
            *    42263    42362: gap of 100 bp
            *    42363    71362: contig of 29000 bp in length
            *    71363    71462: gap of 100 bp
            *    71463    90991: contig of 19529 bp in length
            *    90992    91091: gap of 100 bp
            *    91092   134901: contig of 43810 bp in length
            *   134902   135001: gap of 100 bp
            *   135002   149352: contig of 14351 bp in length
            *   149353   149452: gap of 100 bp
            *   149453   153318: contig of 3866 bp in length
            *   153319   153418: gap of 100 bp
            *   153419   159997: contig of 6579 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159997
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-221H20"
                     /clone_lib="DanioKey"
     misc_feature    1. .38359
                     /note="assembly_fragment:01793
                     fragment_chain:1"
     misc_feature    38460. .42262
                     /note="assembly_fragment:03784
                     fragment_chain:1"
     misc_feature    42363. .71362
                     /note="assembly_fragment:01035
                     fragment_chain:1"
     misc_feature    71463. .90991
                     /note="assembly_fragment:00574
                     fragment_chain:1"
     misc_feature    91092. .134901
                     /note="assembly_fragment:02665
                     fragment_chain:1"
     misc_feature    135002. .149352
                     /note="assembly_fragment:00176
                     fragment_chain:2"
     misc_feature    149453. .153318
                     /note="assembly_fragment:03848
                     fragment_chain:2"
     misc_feature    153419. .159997
                     /note="assembly_fragment:00001.0"
ORIGIN

  Query Match             10.9%;  Score 109.6;  DB 14;  Length 159997;
  Best Local Similarity   46.2%;  Pred. No. 0.042;
  Matches  431;  Conservative    0;  Mismatches  499;  Indels    2;  Gaps    2;

Qy         28 CGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTGAGCGACGAGAGGCG 87
Db      41330 CGCCCGCGCGGCGGGCCCGCCCCCCCGCGCCGCGCGCCGCCCGAGCCCCCGGGGGGGCC 41389

Qy         88 CTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTGGCCATGCACGCGAG 147
Db      41390 GGCGCCCCGGCCGCGCCCCGCCGCGCGCCCCCCGCGCCGGGCGCCGCCCGCCGCCCCGC 41449

Qy        148 CGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGCGCGAAGCCGACGCC 207
Db      41450 CCGCCCGGGGCCCCCGCCGCGGCGGCCGGCCGCCGCGCGCCGCCGCGGGCGCCCGCCCCC 41509

Qy        208 GCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCACGAGCGGGCT 267
Db      41510 CGCCGCCCGCCGCCGCCGCCCGGCCCCGCGCCGGCGCGCGCGGGCGCGCCCGGGGCCGCG 41569

Qy        268 CCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGATCGCGGCGAGCAC 327
Db      41570 CCCCCGGCGGCCGGCGCCGCCGCGCGCGCCCGCCCCGGGGGCGCCGGCGCGCGGGGGCGC 41629

Qy        328 GCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCACGGCGGCGGCTCC 387
Db      41630 GCGGCGCCGCGGCGCCGGCCGCGCCCCCCCGCCGCCGGCGCGGGCGGCGCGCCCCCCGG 41689

Qy        388 CGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGCGCGGCGAGCGCGATC 447
Db      41690 CGCCGCGCCGCGCCCGCCCCGCGGCCGTCCGGCGCGGCGGGCCGCGCCGCCGCCCGCGGC 41749

Qy        448 AGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGCGGGCCCAGAGCAAG 507
Db      41750 CCGCGCCCCCGGCCCGGACCGCGGGCGCCCCCGGCGCGCGGCCCCCCCCGCGGCCGCCGC 41809

Qy        508 CGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCACGAGCGCGCGATCCGCC 567
Db      41810 GCCGCCCGCCCCCGGCGGGGGGGGTGCGCGGCGGCCGGGCGGGGGGGGGCCCGGCCGCCC 41869

Qy        568 GCGCCGGCCTCTGCGTCGCG-TCCTCCGCCCATGGGCAGAGGCTACTCAGGGCCGCCGCG 626
Db      41870 CCGGCCGCGCGGGCGCCGCGCGCCCGCGCCGGAGGGCGGGCGCCGGCGGGGGCCCCGGGC 41929

Qy        627 GCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGTCTCAATCGCAGCCGCGGA 686
Db      41930 GCGGGCCGGGCGCACGGGCCGGGCGGTGTCGCGCGCCGCGCCCCGCTGCCGCCGCCGCGGG 41989

Qy        687 TCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCGCCGATCCGCCTCTTCCGC 746
Db      41990 GCGGCCGGGCGCCGGCAGGGCCCGGGGCCCAGCCCCGGCCGGCGCGCGCGCCCCCCCC 42049

Qy        747 GGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCTCGGCCGCGATCGTGGCGA 806
```

```
Db      42050 CGGCCCCGCTCGCGCCGGGGCGCGGGCATGGCCGGCCCCCCCCACCCCCCGGCCCGGCGCC 42109

Qy        807 GCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCGCGATCTTCTCGAGGTGAC 866

Db      42110 CCGCGCCGCCCCCCCCGGGCCCCGGCGGCGCGGCGGCCGCCGGCCGGCCCCGCCCCGCGTC 42169

Qy        867 TGCGCGCGT-GCTCGATCACGGCCTCGTTGCCCATGTCGATCCCCCACTTCGCCGCGAGC 925

Db      42170 GCCGGCCGCCGCCCCGCCGCGCCGCCGCCCGCGCCGCCCCGGGCCGGCGTCGCCGCGACG 42229

Qy        926 GCGGGCCACGCCGCCCAGCGCTCGGCGGCGTG 957

Db      42230 CGCCGCCCGCCCGCGCGGCGCGCGGCGGCGCG 42261


RESULT 7
AX078627/c
LOCUS       AX078627               1677 bp    DNA     linear   PAT 22-FEB-2001
DEFINITION  Sequence 141 from Patent WO0107624.
ACCESSION   AX078627
VERSION     AX078627.1  GI:13158269
KEYWORDS    .
SOURCE      Pseudomonas putida
  ORGANISM  Pseudomonas putida
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1
  AUTHORS   Fraser,C.M., Venter,C., Tuemmler,B., Hoheisel,J., Duesterhoeft,A.,
            Hilbert,H., Timmis,K.N., Moore,E., Straetz,M., Heim,S.,
            Nelson,K.E., Hickey,E. and Peterson,J.
  TITLE     Dna sequences which are suited for specifically detecting
            Pseudomonas putida kt2440
  JOURNAL   Patent: WO 0107624-A 141 01-FEB-2001;
            THE INSTITUTE FOR GENOMIC RESEARCH (US) ; QIAGEN GmbH (DE) ;
            Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;
            Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische
            Hochschule Hannover (DE)
FEATURES             Location/Qualifiers
     source          1. .1677
                     /organism="Pseudomonas putida"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:303"
ORIGIN

  Query Match            10.7%;  Score 107.4;  DB 6;  Length 1677;
  Best Local Similarity  55.9%;  Pred. No. 0.32;
  Matches  204;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62

Db        901 CAATCAGCATGATGATGGTGGTCAAGGCCCCGGCCCAGGTACTGCGGAAGCGGAAGGCCA 842

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122

Db        841 GCAGGTCGGCCAGCGATGACAGCTGATAGGTGCGGGTGATCTTGAGGATCGGGTAGAGCA 782

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182

Db        781 GCACCGGCGCCAGCAGGAATGCGCCAGACACCCCCAGGTAACAGGCAAGGAAGCCGTAGC 722

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242

Db        721 CGTACTGGTAGGCCAGGCCAACCGAGCCATAGAAGGCCCAGGCACTGGCGTAGACGCCCA 662

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302

Db        661 GCGACAGGGTGTAGGTAAGTGGGTGGCGAATGATCGAACGCGGGATCAGCCCGCGCTCGC 602

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362

Db        601 TGATCCAGGCCACGCCGAACAGCACCATGAGGTACCCGGCGCTGATCAGGATCATGTGGG 542

Qy        363 CCACG 367

Db        541 TCAGG 537


RESULT 8
PM11H12G
LOCUS       PM11H12G               1393 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm11h12.g, sequence tagged site.
ACCESSION   AL684264
VERSION     AL684264.1  GI:19338009
KEYWORDS    STS.
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 1393)
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .1393
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm11h12.g"
ORIGIN

  Query Match            10.7%;  Score 107.2;  DB 10;  Length 1393;
  Best Local Similarity  45.2%;  Pred. No. 0.35;
  Matches  450;  Conservative    0;  Mismatches  530;  Indels   16;  Gaps    4;

Qy          9 GCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGT 68

Db        402 GNGCGCCGCCGNNNCCGCCCGCCCGCGCNNGCCGCCGCCCGGCCGGGCGCGCGCCCGCGC 461

Qy         69 CGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCG 128

Db        462 CGCGCGCGCGCGCGCCGCGCGCGCCGCCCCCGCGGGGGGGNNGGGGCGNNGNCCCGCCGG 521

Qy        129 GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGC 188

Db        522 CCGCCGCCCGCGGCGCGCGCG--CGCCGGAACGCCGCNCCGGNGCCGNGGCGCGGGGGCG 579

Qy        189 GCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACG----CCGAGC 244

Db        580 GCGGCGCGGCGCGGGGGGGCCCGCCGCACGCCGCCGCCCCGCCGCNNCCGGCGCCCGCCC 639

Qy        245 GAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCG 304

Db        640 GCGCCGGGGGCGGCGCGGGCGCCCGGCCCGCCGCGGGCCGCGGGCGGGGCCGGGGGCGGG 699

Qy        305 GCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCC 364

Db        700 GCGGGGNCGGCGGGCGGGGGCGCGGCGCGGCGGGCCGCCGCGNGCGGCGCGGGCGCGNCG 759

Qy        365 ACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCG 424

Db        760 CCGNCCGCCCGCGCCGGGGGGCGCGNNGAGAGGGGGGGGGGNGCGGCGGGCGGCGGCGCG 819

Qy        425 TCGCGAGCGCGGCGAGCGCGATCAGC---GCGAGCCACACCGCGAAGACGGCCGCCACCG 481

Db        820 CCGGCCGCCCGCCCGCGCCGGGCGGCGCGGCGGGCGGGGNCGCGCCGCCGGGCGCCGCCG 879

Qy        482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541

Db        880 CGCGCGCCGCGNGCGCGCGCGCCGCCCCGACGGGACGCCCGNCCCGCCGGCCGGCGCGCG 939
```

```
Qy        542 GCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGG 601
              |     || |   ||||||  |  |    | || |  | |    | ||     ||  |    |
Db        940 GGCGCCGCGCCGNCGCGCGCGCGCCGCGGGCGCCGGCGGGCGCCCGNGGCGCGNGCCGCG 999

Qy        602 GCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTA 661
              | |||  |    || ||| |  || |    ||   |  |    |   |   ||
Db       1000 CGCGCGGCGGCGNGCGGGCGCGGGCGCCGCGCGCGNCGGCGCCGGNGCGCCGCGGCCGCC 1059

Qy        662 TTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCG 721
                  |      | ||| | ||||        || ||| |    |||| |   ||  ||||
Db       1060 CGCCGNCGGNCACCGCGGGCGCG-------GGCGCGCGNNGCGCCGCCGNGCGGGCGCCG 1112

Qy        722 CGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCAC 781
              |  ||    |    || | |   |||  |||| ||| |||| |       ||||||  |
Db       1113 CCGGCGCGCCGNGCGACGCNCGCGCNCCGCGGGCGCGCGCAGCGAGCGACGCGCGGCCGG 1172

Qy        782 CCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTC 841
              ||||  | ||  ||   |   || ||  | |    ||| || |||     |  | || |
Db       1173 ANGCGCGCCGCGTCGGCGGGNCCGCGCGCGNCCGCGGCGGGCCCGCCGCGGCGGCGGCGC 1232

Qy        842 TCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATG 901
              |||  |  |   |    |||     |   ||||| || ||     | | |   |
Db       1233 GCGGACCGGCGCGCGGCGAGCGCGCGCGGCGCGCGCACGGGACCNGNCGAGAGACGGCGN 1292

Qy        902 TCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCC 961
              |     |    |   |  |    ||| |  ||| | |     ||||    | ||
Db       1293 GNGNGAGGNGGCGCGCGCCAGCCACGCCGCCGCCGCGCGCCGGCAGAGGCGAGACGAGCG 1352

Qy        962 GCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCG 997
                  || |||| ||  ||||    ||   | ||
Db       1353 CGACCCGCGCGCCACGCNCGACCANNCGCGCACGCG 1388
```

```
RESULT 9
AY929300/c
LOCUS       AY929300               4957 bp    DNA     linear   BCT 12-JUL-2005
DEFINITION  Pseudomonas putida MhaS (mhaS) and MhaR (mhaR) genes, complete cds.
ACCESSION   AY929300
VERSION     AY929300.1  GI:62860911
KEYWORDS    .
SOURCE      Pseudomonas putida
  ORGANISM  Pseudomonas putida
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 4957)
  AUTHORS   Arias-Barrau,E., Sandoval,A., Naharro,G., Olivera,E.R. and
            Luengo,J.M.
  TITLE     A two-component hydroxylase involved in the assimilation of
            3-hydroxyphenyl acetate in Pseudomonas putida
  JOURNAL   J. Biol. Chem. 280 (28), 26435-26447 (2005)
   PUBMED   15866873
REFERENCE   2  (bases 1 to 4957)
  AUTHORS   Arias-Barrau,E., Luengo,J.M. and Olivera,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2005) Bioquimica Y Biologia Molecular,
            Universidad De Leon, Campus De Vegazana S/N, Leon 24007, Spain
FEATURES             Location/Qualifiers
     source          1. .4957
                     /organism="Pseudomonas putida"
                     /mol_type="genomic DNA"
                     /strain="U"
                     /db_xref="taxon:303"
     gene            421. .3396
                     /gene="mhaS"
     CDS             421. .3396
                     /gene="mhaS"
                     /note="two-component signal sensor"
                     /codon_start=1
                     /transl_table=11
                     /product="MhaS"
                     /protein_id="AAY16575.1"
                     /db_xref="GI:62860912"
                     /translation="MPMSFSLTQMILISAGYLMVLFGVAWISERGLIPRSIIRHPLTY
                     TLSLGVYASAWAFYGSVGLAYQYGYGFLACYLGVSGAFLLAPVLLYPILKITRTYQLS
                     SLADLLAFRFRSTWAGALTTIIMLIGVLPLLALQIQAVADSISILTGEPVKARVAFAF
                     CTLIILFTIFFGSRHIATREKHEGLVFAIAFESVIKLLALGGIGLYALYGVFGGPHEL
                     EVWLLQNQTALAALHTPLQEGPWRTLLLVFFASAIVMPHMYHMAFTENLNPRSLVSAS
                     WGLPLFLLLMSLAVPLVLWAGLRLGATTNPEYFTLGLGIAANNEALALLAYVGGLSAA
                     SGLIIVTTLALSGMALNHLVLPLYQPPAEGNIYRWLKWTRRALIVAIITAGFMFYLTQ
                     NNHQSLANLGIVAFVATLQFLPGVLSVLYWPTANRRGFIAGPLAGTLVWMVTMLLPLL
                     GNLQGFYIPLLDMIYVLDGTSWRMAAIASLAANVLLFTLISLFTNASTEEVSAAEACA
                     VDNVRRPQRRELHAASPQEFATQLAKPLGAKAAQKEVEQALRDLYLPFDERRPYALRR
                     LRDRIEANLSGLMGPSVAQDMVETFLPYKSGNENYVTEDIHFIESRLEDYHSRLTGLA
                     AELDALRRYHRQTLQELPMGVCSLAKDQEILMWNKAMEELTGIAAKHVVGSRLLTIDE
                     PWRGLLQGFINLPDEHLHKQRLALDGQPRWLNLHKPAIDEPLAPGNSGLVLLVEDLTE
                     TQALEDKLVHSERLASIGRLAAGVAHEIGNPITGIACLAQNLREEREGDGEIIELSSQ
                     ILDQTKRVSRIVQSLMSFAHAGGSHQNSEEPVCLAEVAQDAIGLLALNRRNFEVQFFN
                     LCNPDHWAEGDPQRLAQVLINLLSNARDASPPGSAVRVRSEVSEHTVDLIVEDEGSGI
                     PKNIMDRLFEPFFTTKDPGEGTGLGLALVYSIVEEHYGQITIDSPADIERQRGTRIRV
                     TLPRHVVATSPEIRDRREN"
     gene            3401. .4843
                     /gene="mhaR"
     CDS             3401. .4843
                     /gene="mhaR"
                     /note="two-component response regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="MhaR"
                     /protein_id="AAY16576.1"
                     /db_xref="GI:62860913"
                     /translation="MPHILIVEDETIIRSALRRLLERNQYQVSEAGSVQEAQERFSIA
                     TFDLIVSDLRLPGAPGTELIKLGQGTPVLIMTSYASLRSAVDSMKMGAVDYIAKPFDH
                     DEMLQAVARILRDRQNAPAAASAAEPRATNGKAASADKGSAAAANGEIGIIGSCPPMQ
                     DMYSKIRKVAPTDSNVLIQGESGTGKELVARALHNLSRRAKAPMISVNCAAIPETLIE
                     SELFGHEKGAFTGASAGRAGLVEAADGGTLFLDEIGELPLEAQARLLRVLQEGEIRRV
                     GSVQSQKVDVRLIAATHRDLKNLAKAGQFREDLYYRLHVIALKLPALRERGSDVNEIA
                     SAFLARQSARIGRDDLHFSAEAEQAIRHYSWPGNVRELENAVERAVILSESAEISAEL
                     LGIDIELSDLEEDEILNNALVVASAASASHEPTEDLSLEDYFQHFVLEHQDHMTETEL
                     ARKLGVSRKCLWERRQRLGIPRRKSNATSE"
ORIGIN

  Query Match           10.7%;  Score 107.2;  DB 1;  Length 4957;
  Best Local Similarity  57.0%;  Pred. No. 0.23;
  Matches  196;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  || |  || |  ||| ||||||||   || |||  ||||||| ||||| |
Db        802 CGATCAGCATGATGATGGTGGTGAGTGCCCCGGCCCAGGTGCTGCGGAAGCGAAACGCCA 743

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              |||||||||  ||||| || ||  | |    |||  |||||    || |  |       |
Db        742 GCAGGTCGGCCAGCGATGACAGCTGGTAGGTGCGGGTGATCTTGAGGATCGGGTAGAGCA 683

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| |    || | |||||   | |   || |||||    || |||||| ||   |
Db        682 GCACCGGTGCCAGCAGGAACGCGCCGGACACCCCCAGGTAACAGGCAAGGAAGCCGTAGC 623

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              | |   | |  || | ||| ||   ||| ||| | | |||  |    |||||| ||||| |
Db        622 CATATTGGTAGGCCAGGCCCACCGAGCCATAGAAGGCCCAGGCACTGGCGTACACGCCCA 563

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |      |||||||   ||   || ||  ||||||  |    | ||||
Db        562 GCGACAGGGTGTAGGTGAGCGGGTGGCGAATGATCGAACGCGGGATCAACCCCCGCTCGC 503

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTG 346
              |      || |   |   |||||   |||  || |  |  ||
Db        502 TGATCCAGGCCACGCCAAACAGCACCATGAGGTACCCGGCGCTG 459


RESULT 10
AE016791/c
LOCUS       AE016791               300933 bp    DNA     linear   BCT 12-DEC-2002
DEFINITION  Pseudomonas putida KT2440 section 18 of 21 of the complete genome.
ACCESSION   AE016791 AE015451
```

```
VERSION     AE016791.1  GI:26557036
KEYWORDS    .
SOURCE      Pseudomonas putida KT2440
  ORGANISM  Pseudomonas putida KT2440
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 300933)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Complete genome sequence and comparative analysis of the
            metabolically versatile Pseudomonas putida KT2440
  JOURNAL   Environ. Microbiol. 4 (12), 799-808 (2002)
REFERENCE   2  (bases 1 to 300933)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .300933
                     /organism="Pseudomonas putida KT2440"
                     /mol_type="genomic DNA"
                     /strain="KT2440"
                     /db_xref="taxon:160488"
     gene            153. .740
                     /locus_tag="PP4517"
     CDS             153. .740
                     /locus_tag="PP4517"
                     /note="identified by match to PFAM protein family HMM
                     PF03695"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAN70091.1"
                     /db_xref="GI:24986250"
                     /translation="MSFAEQLTRLQAFLDADELHEEALDYVAAHGYLTALSINAEDVP
                     EREWIDALFAEEPHYASEAQRTEIEATLVALKAHIARQLASDEEFDLPCDLDLTDEPD
                     DSDLRGWCIGFMEGVFLREEAWFENAEEEVSEMLLPIMVGSGLFDEQPEFADIASNAN
                     LQDDMIVQIPEALSALFLLLHAPDEKPALLKPRHH"
     gene            748. .1113
                     /locus_tag="PP4518"
     CDS             748. .1113
                     /locus_tag="PP4518"
                     /note="identified by match to PFAM protein family HMM
                     PF04304"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAN70092.1"
                     /db_xref="GI:24986251"
                     /translation="MRYLLLATGWLSVALGVLGIFLPVLPTTPFLLLAAACFARSSPR
                     FHHWLVHHPKLGPWIRDYLSGEGIPLKGKVYAIGLMWASIGLSCYLVPLFWARTFMLT
                     SAALVTLYILSLKTLHKPK"
     gene            1425. .2783
                     /gene="tolC"
                     /locus_tag="PP4519"
     CDS             1425. .2783
                     /gene="tolC"
                     /locus_tag="PP4519"
                     /note="similar to GB:X12387, SP:P05184, SP:P08684,
                     SP:P20815, PID:181354, PID:181372, PID:181374, PID:219963,
                     PID:35911, and PID:510086; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="agglutination protein"
                     /protein_id="AAN70093.1"
                     /db_xref="GI:24986253"
                     /translation="MRVLNPITSALLLALASANVQAMSITEAVQSAVDYHPQVSSNRN
                     SKLSADEDVKFARGGYYPSVDLVAGYGRQRSDNATTRAEGNHNKETLNYTQSELRLRQ
                     MIFDGFNTSNEVGRTEAVSTSRAYYTQAVAQDVALRAVEVYLEVLKRRELVTLAKNNL
                     QAHLRVNDQIGLRNERGVGSTADLDQSRARRALAENNLDTAEVDLADAEANFFSVIGR
                     APDELESPQSIKGEVPGTLEEARDGMRQNNPYIKSAQADVNAAEKQYEVGKSTFYPRF
                     DAILATGANNNTGGEKGHNNNDWQAGVEMNYNLFRGGSDKARLQSDAHKINQALDIRN
                     NALRELTENLSLAWNAMNNASKQLPTAREYAETTKRVRAAYQDQFGLGQRTLLDVLDS
                     ENELYNADRRYTEVRYTEEFSRYRVLATMGELLSKQHISLPPEALATTEVRTEARLPE
                     MR"
     gene            complement(2846. .3445)
                     /locus_tag="PP4520"
     CDS             complement(2846. .3445)
                     /locus_tag="PP4520"
                     /note="identified by match to PFAM protein family HMM
                     PF03942"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAN70094.1"
                     /db_xref="GI:24986254"
                     /translation="MPRPRCERCQRPLDHCLCPLIPALDSRTRVILLQHPSETAHALN
                     TARLAALGLNNAELRIGEVFDDLHDLLATPGYHPVLLFPGDDAQVLTAYGEAGDKPLL
                     LIVPDGTWRKARKLLYLNPLLGGLPRVTLGQVAASRYRLRKAPEAGALSTIEAVVGAL
                     NTLEQPACFDALLAPFETLIEGQIRAMGVETFQRNHGRE"
     gene            complement(3510. .5075)
                     /locus_tag="PP4521"
     CDS             complement(3510. .5075)
                     /locus_tag="PP4521"
                     /note="similar to GP:9947522; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="aerotaxis receptor, putative"
                     /protein_id="AAN70095.1"
                     /db_xref="GI:24986255"
                     /translation="MRMNLPVTEHERTFPGDQRLISTTDLDSRITYCNDAFVAISGFT
                     YDELVGQPHNLVRHPDMPPAVFGHMWETIKQGKPWMGIVKNRAKNGDYYWVSAYVTAI
                     YEQGRISGYESVRSVPTREQIRRAEALYARLRDGRSPVPWAARLGHGLGHGWPLIGAG
                     LLSAAGYLWLPPYAALAVLMASLLLAWYWVEHRQNQAIRRTLAEHPKAFTSPLVALTY
                     SDNPGLKGQLDLAIISEEARLQTALTRLVDAGVGVKSRAAQSSDLSDAQAQMLDRQRS
                     ETDQSATAIAQMAATIQQVTHNVQSTAHAASDADQLAQQGSELALKSLKDMGSMSDAV
                     NDIGQAVNALAEQTQSIGSVVDVITSIAEQTNLLALNAAIEAARAGEQGRGFAVVADE
                     VRSLAQRTRASTEEIHHIIASLRAGAERAVSTASRGEQISRDSVHSVEAVQAALSGIA
                     VAVSRITGMSQQMATASEQQSHVAEDINQQIVRIAQLCDQSAGQARQGAEISQDLERM
                     AEYLHSLAERFNR"
     gene            complement(5214. .6107)
                     /locus_tag="PP4522"
     CDS             complement(5214. .6107)
                     /locus_tag="PP4522"
                     /note="similar to GP:9947370; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, LysR family"
                     /protein_id="AAN70096.1"
                     /db_xref="GI:24986256"
                     /translation="MASTLPDLKLLRIFISVVRHQGFANAQRELNLSTSAISTYMSQL
                     EGALGIVLCHRGRGGFSLTSKGELFHQETLRLLAELDGFEQYAAALKGELRGTLNLGV
                     IDSTVGDRALPLAEAIGAYSQEHPAVHLHLSVSSPYELQLGVQDNRLDLAIGAFSSRM
                     SGLLYQPLYREQHWLYCSSRHPLYAERRIPEQVVTQQRMVGRGYWSQAELARHGFKHS
                     AATVESMEAQLILILSGAYIGYLPEHYAQAWVDKGDLRVLSPSTFGYQAPFSLIIRRG
                     RSREPLIQTFRDLLKSQLNVG"
     gene            6320. .7282
                     /locus_tag="PP4523"
     CDS             6320. .7282
                     /locus_tag="PP4523"
```

```
                     /note="similar to SP:P37819; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="agmatinase, putative"
                     /protein_id="AAN70097.1"
                     /db_xref="GI:24986257"
                     /translation="MRPTVDKTLHQPLGGNEMPRFGGIATMLRLPHLQSAKGLDAAFI
                     GVPLDIGTSLRSGTRFGPRQIRAESVMIRPYNMATGAAPFDSLSVADIGDVAINTFNL
                     LDAVRIIEEAYDEIVEHNVIPMTLGGDHTITLPILRALHKKHGKIGLVHIDAHADVND
                     HMFGEKIAHGTTFRRAVEEGLLDCDRVVQIGLRAQGYTADDFNWSRRQGFRVVQAEEC
                     WHKSLEPLMAEVREKVGGGPVYLSFDIDGIDPAWAPGTGTPEIGGLTTIQAMEIIRGC
                     HGLDLIGCDLVEVSPPYDTTGNTSLLGANLLFEMLCVLPGVVRR"
     gene            7415. .8794
                     /locus_tag="PP4524"
     CDS             7415. .8794
                     /locus_tag="PP4524"
                     /note="identified by match to PFAM protein family HMM
                     PF04279"
                     /codon_start=1
                     /transl_table=11
                     /product="sodium-solute symporter, putative"
                     /protein_id="AAN70098.1"
                     /db_xref="GI:24986258"
                     /translation="MALDIIVVMIYTAGMLGLGWYGMRRAKTHEDYLVAGRNLGPVLY
                     MGTMATTVLGGASTVGTVRLGYVHGISGFWLCAALGLGIIAINLFLAKPLLRLRIFTV
                     TQVLEQRYNPTARQASAVIMLAYALMIGVTSTLAMATVLQVLLDLPFWASLMLGGGVV
                     VLYSTIGGMWSLTLTDIVQFVIKTVGLMFILLPVCLYKAGGWDTLVAKLPAASFQLTT
                     IGWDTIITYFLIYFFGILIGQDIWQRVFTARDEKVCQRAGTAAGVYCVLYGLACAAIG
                     MAAHVLMPDLANPNNAFAEMIKTTLPDGIRGLLMAAALAAMMSTASAGLLAASTTVTE
                     DLLPKLRGGKQSSLGTSRLFTLLTGLVVLGIALMVNDVINALTLAYNLLVGGMLIPLI
                     GAIFWKRATTAGAIASMSLGFATALLFMFLDGLEANTPIYYSLAVGLVSFVGVSLVSR
                     KSVAAVKLA"
     gene            complement(9174. .10067)
                     /locus_tag="PP4525"
     CDS             complement(9174. .10067)
                     /locus_tag="PP4525"
                     /note="similar to GP:9947361; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, LysR family"
                     /protein_id="AAN70099.1"
                     /db_xref="GI:24986259"
                     /translation="MEFSQLRIFQAVAEEGSVTRAAERLHRVPSNLSTRLRQLEEQLG
                     VELFLRERQRLQLSPAGKVLLDYANRMSALRDEALGAVRGGQPAGDFVLGTMYSTAAT
                     HLPALLARYHQSYPQVNLQVRAAPSGELLEGLLNHTLDAALVDGPPSLAGLDGVPLCD
                     EQLVLITSPEHPAVLTAKDVAGKAVFTFRQGCSYRMRLEAWYAHAHTPMGRVMEIESY
                     QSMLACVIAGAGVAMMAQSMLDSLPGRDRVRVHSLQAPFDQAVTWLMWRQGMRGANLQ
                     AWIDLQQSETINQSSECAATA"
     gene            10172. .11344

  Query Match             10.7%;  Score 107.2;  DB 1;  Length 300933;
  Best Local Similarity   57.0%;  Pred. No. 0.055;
  Matches  196;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  || |  || |   |  ||||||||   ||  ||  |||||||||| || |
Db     202458 CAATCAGCATGATGATGGTGGTCAAGGCCCCGGCCCAGGTACTGCGGAAGCGGAAGGCCA 202399

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              |||||||||  ||||| || ||  | |    |||  |||||    || |   |        |
Db     202398 GCAGGTCGGCCAGCGATGACAGCTGATAGGTGCGGGTGATCTTGAGGATCGGGTAGAGCA 202339

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | |||||    || | | |||   | |   || |||||    || |||||| ||   |
Db     202338 GCACCGGCGCCAGCAGGAATGCGCCAGACACCCCCAGGTAACAGGCAAGGAAGCCGTAGC 202279

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   ||| ||| | | ||| |    ||||||||||||  |
Db     202278 CGTACTGGTAGGCCAGGCCAACCGAGCCATAGAAGGCCCAGGCACTGGCGTAGACGCCCA 202219

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |      || |||   ||   || ||  ||||||  | |  ||||||
Db     202218 GCGACAGGGTGTAGGTAAGTGGGTGGCGAATGATCGAACGCGGGATCAGCCCGCGCTCGC 202159

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTG 346
              |       || |   ||   |||||   |||  || |  |  ||
Db     202158 TGATCCAGGCCACGCCGAACAGCACCATGAGGTACCCGGCGCTG 202115


RESULT 11
PM2D12B
LOCUS       PM2D12B                  956 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm2d12.b, sequence tagged site.
ACCESSION   AL684743
VERSION     AL684743.1  GI:19337611
KEYWORDS    STS.
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 956)
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm2d12.b"
ORIGIN

  Query Match             10.5%;  Score 105;  DB 10;  Length 956;
  Best Local Similarity   46.2%;  Pred. No. 0.63;
  Matches  340;  Conservative    0;  Mismatches  392;  Indels    4;  Gaps    3;

Qy          7 GAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAG 66
              | | | |  || ||  | |||  |||||||    ||||  |    ||  |  |    |  ||
Db        166 GGGGGGGGGGGGGGGGGGGAGGGCCCCGCGCGCGGNGNNGGGGGGGGNGNGGGGCCGCNC 225

Qy         67 GTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAG 126
              | |||   | |  |   |  |  |  |||| | |  |||  || |||  | || | ||
Db        226 GCCGGNCGGGGGGGNNNGNNGGNGGGCGCGGGGGGCGCGGCGCGCGCCGGGCCCGCGGGC 285

Qy        127 CGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTG 186
              | |  |  |  |  |  | | | | |  |  | |      | ||| ||    || |     |
Db        286 CCCCCGGCNCCNGGGGGGGGGGGGCCCGGGGGGCGGGCCGGCGGGGGGNNCGGGGGGCGG 345

Qy        187 GCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGA 246
              |  || |      || | | | ||  |  || || ||    |||| | ||| |
Db        346 CGGGGGCCCCNCCGCGCCCCGGGCG--GNCCGGCCGCGNNCCGGCGCNGGCGGGGGCGNC 403

Qy        247 GAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGC 306
              | | ||   |  | |||  |     | ||||| | | | | |||||||| ||  ||
Db        404 GCGCGCCCNCCGGGGCGNNCNGGCGCGGGGCGCGGGGGCGGGGCGGGCGCGCGGCCGGGG 463

Qy        307 GAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCAC 366
              |  |  |||||   |  ||||| |||  | |   |  | |    || || | ||| |
Db        464 NGCCCGGNGCGCGGGCGGGGGGCCGCGCCGGGGCGGGGGCGNGGNNGGGGGGGCCCCCGC 523

Qy        367 GTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCG-GCGCGCGT 425
              | ||       |||| |   | || ||   ||  |||||  | || |  || ||| |||
Db        524 GGCGCGNGNNGCGGGGCGCCGGGCGCGGGCCGCGCGGCGCGGGCCGGCGCGCGCGGGCGG 583

Qy        426 CGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAG 485
              |||| |  ||        |  | | |  ||      |   ||| || | ||||| |
Db        584 CGCGGGGCGGGGCGCGNGGGNCGGGGGCGGCGGGGGGCGGCGCGCGGGCGGCGCCGCGGG 643
```

```
Qy        486 CGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCAC 545
              || | |||| |      ||  ||||||        | ||  |    | | |     |
Db        644 CGCGNCGCGCGGGGCCGGCCGGCGCGCGNNNNCGGCCCGGGCGCGGCCCCNCGNGGCCCG 703

Qy        546 GAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAG 605
              | || |   |||||     ||  | |    |        |||  | |||||     |  |
Db        704 GCGCCCCGNCGCGCCGNGNGCGCCCCGCNNCGNGCGCNNGCGGACGCCGCCGGGCCGGNG 763

Qy        606 AGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGT 665
              | |  | | | || || | ||  |     | | |       | | |   | |     |
Db        764 CGCCGCCGCCGCGCGGCGGGGGGCGNCGGGGCGCGCGGCCCGGGCCCGGCGGGGGGGNGC 823

Qy        666 GGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGC-CGCGGATGGGCAGCCGCGA 724
              ||         |   | |||   ||  | |||||    || || |||  ||  || |||
Db        824 GGCGCGGGCGGNGCCGCCGGCCGGGGCGCGCGGGCGGGGCGCGNGGACCGGGGGCGGCGG 883

Qy        725 GCCGCCGATCCGCCTC 740
              |||| ||  ||| | |
Db        884 GCCGGCGCGCCGGCGC 899


RESULT 12
CR392332/c
LOCUS       CR392332              207592 bp    DNA     linear   HTG 19-APR-2005
DEFINITION  Danio rerio clone DKEY-58D5, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   CR392332
VERSION     CR392332.3  GI:62750901
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 207592)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 19, 2005 this sequence version replaced gi:50950048.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK58D5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 205163 bases at least Q40
            Consensus quality: 205691 bases at least Q30
            Consensus quality: 206076 bases at least Q20
            Insert size: 207192; sum-of-contigs
            Insert size: 198347; 3.6% error; agarose-fp
            Quality coverage: 7.02x in Q20 bases; sum-of-contigs Quality
            coverage: 7.44x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    23007: contig of 23007 bp in length
            *    23008    23107: gap of 100 bp
            *    23108    51856: contig of 28749 bp in length
            *    51857    51956: gap of 100 bp
            *    51957   102632: contig of 50676 bp in length
            *   102633   102732: gap of 100 bp
            *   102733   172682: contig of 69950 bp in length
            *   172683   172782: gap of 100 bp
            *   172783   207592: contig of 34810 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207592
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-58D5"
                     /clone_lib="DanioKey"
     misc_feature    1. .23007
                     /note="assembly_fragment:02417
                     fragment_chain:1"
     misc_feature    23108. .51856
                     /note="assembly_fragment:00001
                     fragment_chain:1"
     misc_feature    51957. .102632
                     /note="assembly_fragment:00572
                     fragment_chain:1"
     misc_feature    102733. .172682
                     /note="assembly_fragment:01654
                     fragment_chain:1"
     misc_feature    172783. .207592
                     /note="assembly_fragment:00260.0"
ORIGIN

  Query Match             10.4%;  Score 104.4;  DB 14;  Length 207592;
  Best Local Similarity   41.7%;  Pred. No. 0.11;
  Matches  386;  Conservative    0;  Mismatches  533;  Indels    6;  Gaps    2;

Qy          7 GAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAG 66
              |||
Db     172785 GAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 172726

Qy         67 GTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAG 126
                                                          ||||||  |   ||
Db     172725 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCGCGCGGGCGCCGGGC 172666

Qy        127 CGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTG 186
               || | |||   || ||   ||| | | ||| ||    | |||  |   | || | | | |
Db     172665 GGGTGCGGCGCGGCGCGGCCGGCGGC-GACCGGGGCGGCCCGCGCGCGCCGCGGCGGCG 172607

Qy        187 GCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGA 246
              |||    |||  | || ||| |       |      |   |  | |  | |||||   |
Db     172606 GCGGGCGGCGGCGGCGGCGCGCCGCGCCGCCCCGGCCCGNGCCGGGCGGGCGCCGCTGGG 172547

Qy        247 GAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGC 306
              | || | ||   | ||      | |   | | ||| | | |   ||||| |||  |||||
Db     172546 GCGGCCTAGGCGGGGCCCCGGCGGGCGGGGGGCGGGGGCGACGCGGGCCCGCCCCGCGGG 172487

Qy        307 GAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCAC 366
              | | |   || ||| |   |||   ||  |  ||| |||| | |     ||||  |
Db     172486 CGGGGGGGCGGCCGCGGGGCGGCCCGCCCCGCCGGCCGCGGCGCGCGG-----CGCCGGC 172432

Qy        367 GTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTC 426
              | ||      || || ||  |||| ||   |||      | |||||| | ||||  |
Db     172431 GCCGGGGGGCGCCGCCGCCGCCGCGGGCGCGCGCCCCGCGCCCCCGGGCCGCGCCGGGG 172372

Qy        427 GCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGC 486
              ||| |   || | | ||     ||| | | |    |||        ||  ||  |||| ||
Db     172371 GCGCGGCGGGGGGGGGCCGGGCGCGGGGGGCCGCGCGCCCGCTGCGCGCCCCGCGCGCGC 172312

Qy        487 GGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACG 546
                ||   ||||  | ||    | ||  |||   |  | |  |     || ||
Db     172311 CCGCGCGGGGCGGGGGGCCGCGGGGCGCGCGCGCGGGCGCGCGCGAGCGCCCGCGCGGGC 172252

Qy        547 AGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGA 606
                | | | | |  | | || |  || | |  | || ||  | |    |   ||
Db     172251 GCGCCCGGGGGGGGGGCGGGCGGGGGGGTCCCCCCCTCCCGCGCGCGCGCGCGCGCGA 172192

Qy        607 GGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTG 666
              |       |||  | ||||   |   || | |  ||   |       |||     |
```

```
Db     172191 GCGCGGGGGGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 172132

Qy        667 GTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGC 726

Db     172131 CTCGCGCGCGCGCGGGGGGGGGGGGGGCCCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGC 172072

Qy        727 CGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCG 786

Db     172071 GCGCGCGCGCGCGCGCGCGCGGGGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 172012

Qy        787 CTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGC 846

Db     172011 CGCGCGCGCGCGCGGGGCGGGTGTGTGTGTGTNTGTCTCTNTNTCTNTNTNTCTGTGTGT 171952

Qy        847 CGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGAT 906

Db     171951 GGGGGGGGGGGGGGGGGGGGGAGGTGTGTGCGCCCCCGCGCGCGCGCGCGCGCGCGCGCG 171892

Qy        907 CCCCCACTTCGCCGCGAGCGCGGGC 931

Db     171891 CGCGCGCGCTCGCGCGCGCGCGCGC 171867


RESULT 13
AE016853_10
WPCOMMENT
  Sequence split into 64 fragments  LOCUS AE016853 Accession AE016853
     Fragment Name      Begin     End
     AE016853_00            1    110000
     AE016853_01       100001    210000
     AE016853_02       200001    310000
     AE016853_03       300001    410000
     AE016853_04       400001    510000
     AE016853_05       500001    610000
     AE016853_06       600001    710000
     AE016853_07       700001    810000
     AE016853_08       800001    910000
     AE016853_09       900001   1010000
     AE016853_10      1000001   1110000
     AE016853_11      1100001   1210000
     AE016853_12      1200001   1310000
     AE016853_13      1300001   1410000
     AE016853_14      1400001   1510000
     AE016853_15      1500001   1610000
     AE016853_16      1600001   1710000
     AE016853_17      1700001   1810000
     AE016853_18      1800001   1910000
     AE016853_19      1900001   2010000
     AE016853_20      2000001   2110000
     AE016853_21      2100001   2210000
     AE016853_22      2200001   2310000
     AE016853_23      2300001   2410000
     AE016853_24      2400001   2510000
     AE016853_25      2500001   2610000
     AE016853_26      2600001   2710000
     AE016853_27      2700001   2810000
     AE016853_28      2800001   2910000
     AE016853_29      2900001   3010000
     AE016853_30      3000001   3110000
     AE016853_31      3100001   3210000
     AE016853_32      3200001   3310000
     AE016853_33      3300001   3410000
     AE016853_34      3400001   3510000
     AE016853_35      3500001   3610000
     AE016853_36      3600001   3710000
     AE016853_37      3700001   3810000
     AE016853_38      3800001   3910000
     AE016853_39      3900001   4010000
     AE016853_40      4000001   4110000
     AE016853_41      4100001   4210000
     AE016853_42      4200001   4310000
     AE016853_43      4300001   4410000
     AE016853_44      4400001   4510000
     AE016853_45      4500001   4610000
     AE016853_46      4600001   4710000
     AE016853_47      4700001   4810000
     AE016853_48      4800001   4910000
     AE016853_49      4900001   5010000
     AE016853_50      5000001   5110000
     AE016853_51      5100001   5210000
     AE016853_52      5200001   5310000
     AE016853_53      5300001   5410000
     AE016853_54      5400001   5510000
     AE016853_55      5500001   5610000
     AE016853_56      5600001   5710000
     AE016853_57      5700001   5810000
     AE016853_58      5800001   5910000
     AE016853_59      5900001   6010000
     AE016853_60      6000001   6110000
     AE016853_61      6100001   6210000
     AE016853_62      6200001   6310000
     AE016853_63      6300001   6397126
Continuation (11 of 64) of AE016853 from base 1000001 (AE016853 Pseudomonas syringae pv.

  Query Match           10.3%;  Score 102.8;  DB 1;  Length 110000;
  Best Local Similarity  55.6%;  Pred. No. 0.19;
  Matches  197;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62

Db      51182 CGATCAGCATGAACACCGTGGTCAGCGCGCCGGCCCAGGTGCTGCGAAAGCGGAAGGCGA 51241

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122

Db      51242 ACAGGTCGGCCAGCGATGATAACTGATAGGTGCGGGTAATTTTCAGGATCGGGTATAGCA 51301

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182

Db      51302 GCACTGGCGCGAGCAGAAACGCCCCGGAAACCCCGAGGTAGCTGGACAGAAAGCCGTAGC 51361

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242

Db      51362 CATACTGGTAGGCCAGACCGACCGTGCCGTAGAACGCCCAAGCGCTGGCGTATACACCGA 51421

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302

Db      51422 GCGACAGGGTGTAGGTCAGAGGGTGGCGAATGATCCAGCGCGGAATCAGCCCGCGCTCGC 51481

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGA 356

Db      51482 TGATCCAGGCCACACCGAACAGCACCAGCAGATAAGCGGCGCTGACCAGCAGCA 51535


RESULT 14
PM2B12B/c
LOCUS       PM2B12B                 1065 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm2b12.b, sequence tagged site.
ACCESSION   AL684695
VERSION     AL684695.1  GI:19337599
KEYWORDS    STS.
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 1065)
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
```

```
     source          1. .1065
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm2b12.b"
ORIGIN

  Query Match           10.1%;  Score 101.4;  DB 10;  Length 1065;
  Best Local Similarity   42.9%;  Pred. No. 1.3;
  Matches  375;  Conservative    0;  Mismatches  495;  Indels    4;  Gaps    2;

Qy        114 GCGCCCAGAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGA 173
Db       1015 GCGNCCGAGCGCGGGNNGNGGGGCGGCGCGCGCGCGCGGGCGCCCGGNCNCCGCGCCGTT 956

Qy        174 AGACGAGCCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGT 233
Db        955 TGCGCNNCCCCGCGCCCCGGCCGNCCCGCGTGGCGCGCCNCGTTGCNCCGCGCGCNCCNC 896

Qy        234 AGACGCCGAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGG 293
Db        895 GCCCCCCCCGCNNCCGNNCCCCGCCGCGCGCGCCGGCCGCGGCGNCCCCGGCGGCNCGCC 836

Qy        294 CGCGCTGCGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACA 353
Db        835 GCCGCGCGGCCCNGCCNGGGCCCGCCGGGCGCNCCGCGGGCCNCCGCNGCGCGGNGCGCC 776

Qy        354 GGAGGACGCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCG 413
Db        775 NCGCNCCGGCGGCGGCGCCNGCCG-GCCCCCGCCGGCGCCCGGCNCCGCCCGCGGCNNCC 717

Qy        414 GTCGGCGCGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGC 473
Db        716 CGCGCCGCGCGCGGCCGCGCGNGCCCCCGGCGCGGGGCGGCCCCCGCCCCNNNNNCGGG 657

Qy        474 CGCCACCGCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGC 533
Db        656 CGCGCGCACGGGCGGCCCGCGCCCGCGCGCGCGCCGGCCCCCGGCCGCGGCCCGGCGGGC 597

Qy        534 GCCCAGCAGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCC 593
Db        596 CGCCCGCCCGCCCCCGGNCGCGGCGTCCGCCGGCGCGCCGCCCGCCGACGGGGGCCCGCG 537

Qy        594 GCCCATGGGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCA 653
Db        536 CCCGCGGGCGCGGCGCGGCCNCCGGCCGGCCCNNCGCGNNGCCCCCNCGCGCCCNGCNCC 477

Qy        654 ATGCGTTATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATG 713
Db        476 CGGGCGCCGCGGCGCCCCCCCGCCGCCCGCGGCGGGGGCGNGGCGNCGCGGCCGGCCCNG 417

Qy        714 GGCAGCCGCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGC 773
Db        416 CGCNGGCNNNGGCCGCGNCCNNGGC---CGCGGGGGGGNGGGGGCNNCGGGGGNGCGGGC 360

Qy        774 GCGATCACCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATC 833
Db        359 CCGGGCGCCCCCCCCCCCGCGGGCCCCCCGCNGGGGGCGGCGGGCGGGGGCGCCCCGGCGG 300

Qy        834 ACCCGCTCTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGT 893
Db        299 CCCCCCCGCCGGCNCCCNNNNCNCCCCCGCGGCNCCGCNNCGCGGCCCCGGCGCNCGCCC 240

Qy        894 TGCCCATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGG 953
Db        239 CCCNCCCGCCCCCCNCNGCCGCGGCCGCGCCCCGCGCNCCCNCCCCCCGGCCCCCGGGGN 180

Qy        954 CGTGGGCCGCGAGGCCGGGCCATGCCGGACCGCC 987
Db        179 NCCCGGCCCCCCCCCCCCCCCCCCCNCCCCCCCC 146

RESULT 15

PM11H12G/c
LOCUS       PM11H12G                1393 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm11h12.g, sequence tagged site.
ACCESSION   AL684264
VERSION     AL684264.1  GI:19338009
KEYWORDS    STS.
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 1393)
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .1393
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm11h12.g"
ORIGIN

  Query Match             9.9%;  Score 99.4;  DB 10;  Length 1393;
  Best Local Similarity   45.2%;  Pred. No. 1.7;
  Matches  448;  Conservative    0;  Mismatches  531;  Indels   13;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
Db       1305 CGCCNCCTCNCNCNCGCCGTCTCTCGNCNGGTCCCGTGCGCGCGCCGCGCGCGCTCGCCG 1246

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
Db       1245 CGCGCCGGTCCGCGCGCCGCCGCCGCGGCGGGCCCGCCGCGGNCGCGCGCGGNCCCGCCG 1186

Qy        124 GAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCC 183
Db       1185 ACGCGGCGCGCNTCCGGCCGCGCGTCGCTCGCTGCGCGCGCCCGCGGNG----CGCGNGC 1130

Qy        184 GTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAG 243
Db       1129 GTCGCNCGGCGCGCCGGCGGCGCCCGCNCGGCGGCGCNNCGCGCGCCCGCGCCCGCGGTG 1070

Qy        244 CGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGC 303
Db       1069 NCCGNCGGCGGGCGGCCGCGGCGCNCCGGCGCCGNCGCGCGCGGCGCCCGCGCCCGCNCG 1010

Qy        304 GGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCC 363
Db       1009 CCGCCGCGCGCGCGGCNCGCGCCNCGGGCGCCCGCCGGCGCCCGCGGCGCGCGCGCNCG 950

Qy        364 CACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGC 423
Db        949 GCGCGGCGCCCGCGCGCCGGCCGGGNCGGGCGTCCCGTCGGGGCGGCGCGCGCGCNC 890

Qy        424 GTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCG 483
Db        889 GCGGCGCGCGCGGCGGCGCCCGGCGGCGCGNCCCCGCCCGCCGCGCCGCCCGGCGCGGGC 830

Qy        484 AGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGC 543
Db        829 GGGCGGCCGGCGCGCCGCCGCCCGCCGCNCCCCCCCCCCTCTCNNCGCGCCCCCCGGCGC 770

Qy        544 ACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGC 603
Db        769 GGGCGGNCGGCGNCGCG----CCCGCGCCGCNCGCGGCGGCCCGCCGCGCCGCGCCCCG 714
```

```
Qy        604 AGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATT 663
              | |  | | |  |||||| | |        |     |  ||       ||   |||
Db        713 CCCGCCGNCCCCGCCCCGCCCCCGGCCCCGCCCGCGGCCCGCGGCGGGCCGGGCGCCCGC 654

Qy        664 GTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCG 723
              |  | | |   ||| | || |   |||  | |  ||  | || |||   ||   ||| |
Db        653 GCCGCCCCCGGCGCGGGCGGGCGCCGGNNGCGGCGGGGCGGCGGCGTGCGGCGGGCCCCC 594

Qy        724 AGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCAC-C 782
              |||||  ||||| |  ||||| | ||     |    || |  |  |||||  | | |
Db        593 CCGCGCCGCGCCGCCGCCCCCGCGCCNCGGCNCCGGNGCGGCGTTCCGGCGCGCGCGCGC 534

Qy        783 CGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCT 842
              ||||  |||| ||   || | |           | | |  ||| | |||     |    |
Db        533 CGCGGGCGGCGGCCGGCGGGNCNNCGCCCCNNCCCCCCCGCGGGGGCGGCGCGCGCGGCG 474

Qy        843 CGGCCGCGATCTTCTCGAG-GTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATG 901
              ||  ||||  |  | || | | |    ||  || ||     |  |  |  |  | | |
Db        473 CGCGCGCGCGCGGCGCGGGCGCGCGCCCGGCCGGGCGGCGGCNNGCGCGGGCGGGCGGNN 414

Qy        902 TCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCA---GCGCTCGGCGGCGTGG 958
              ||     | | |  ||| |     |||  || ||||||  |    ||||    | ||| |
Db        413 NCGGCGGCGCNCGGCGCGGGCGCGGCGCNCCNCGCCGCGCNNCCGCGCNGCCCCGCGCGC 354

Qy        959 GCCGCGAGGCCGGGCCATGCCGGACCGCCGGC 990
              ||||  || |||||    ||  | |||  |
Db        353 CGCGCGCCGCGGGGCCNGCGCGCCCGGCCNCC 322
```

Search completed: April 28, 2006, 02:39:32
Job time : 5548.62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 482.694 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 cgccacgagcgcgaaggcgg.....accgccggccgcctcgaagc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1001	100.0	76994	12	ADP64454	Adp64454 Sorangium
	2	137.6	13.7	9867	14	ACL64545	Acl64545 M. xanthu
c	3	135.8	13.6	2973	14	ACL73103	Acl73103 M. xanthu
c	4	123.4	12.3	1172	6	ABQ31996	Abq31996 Oligonucl
	5	123.4	12.3	1172	6	ABQ31997	Abq31997 Oligonucl
c	6	122.4	12.2	1172	6	ABQ31995	Abq31995 Oligonucl
	7	122.4	12.2	1172	6	ABQ31994	Abq31994 Oligonucl
	8	122.2	12.2	1172	6	ABQ31995	Abq31995 Oligonucl
	9	122.2	12.2	1172	6	ABQ31996	Abq31996 Oligonucl
c	10	122.2	12.2	1172	6	ABQ31997	Abq31997 Oligonucl
c	11	122.2	12.2	1172	6	ABQ31994	Abq31994 Oligonucl
	12	113.8	11.4	2217	11	ABD12116	Abd12116 Pseudomon
c	13	113.8	11.4	2526	11	ABD11744	Abd11744 Pseudomon
c	14	113.8	11.4	2991	11	ABD11843	Abd11843 Pseudomon
c	15	107.8	10.8	114955	2	AAX53491	Aax53491 Human ade
c	16	107.4	10.7	1677	4	AAF61095	Aaf61095 P. putida
	17	107.2	10.7	114955	2	AAX53491	Aax53491 Human ade
c	18	104.6	10.4	1951	6	ABQ19256	Abq19256 Oligonucl
	19	104.6	10.4	1951	6	ABQ19257	Abq19257 Oligonucl
c	20	102	10.2	1951	6	ABQ19255	Abq19255 Oligonucl
	21	102	10.2	1951	6	ABQ19254	Abq19254 Oligonucl
c	22	93	9.3	9975	8	AAL61173	Aal61173 Actinosyn
	23	93	9.3	82746	8	AAL61224	Aal61224 Actinosyn
	24	92.8	9.3	3024	14	AEB57391	Aeb57391 Silk elas
	25	90.4	9.0	1951	6	ABQ19256	Abq19256 Oligonucl
c	26	90.4	9.0	1951	6	ABQ19257	Abq19257 Oligonucl
c	27	87.4	8.7	5223	8	ACA26233	Aca26233 Prokaryot
c	28	87.2	8.7	20394	4	AAF24892	Aaf24892 Pimaricin
c	29	87	8.7	88421	6	AAL40781	Aal40781 88421nt g
	30	86.4	8.6	1951	6	ABQ19255	Abq19255 Oligonucl
c	31	86.4	8.6	1951	6	ABQ19254	Abq19254 Oligonucl
c	32	84.8	8.5	2118	8	ACA26211	Aca26211 Prokaryot
	33	84.6	8.5	14055	8	AAL61170	Aal61170 Actinosyn
c	34	84.6	8.5	82746	8	AAL61224	Aal61224 Actinosyn
c	35	84	8.4	14055	8	AAL61170	Aal61170 Actinosyn
c	36	83.4	8.3	1371	8	ACA26384	Aca26384 Prokaryot
c	37	83.4	8.3	29879	2	AAQ46806	Aaq46806 eryA regi
	38	83	8.3	12425	5	AAH26495	Aah26495 Human low
	39	81.4	8.1	1614	5	AAH26499	Aah26499 Human low
c	40	81	8.1	884	8	ACA26257	Aca26257 Prokaryot
c	41	80.6	8.1	125401	4	AAD17186	Aad17186 Streptomy
c	42	79.8	8.0	3024	14	AEB57391	Aeb57391 Silk elas
c	43	79.6	8.0	1119	8	ACA26139	Aca26139 Prokaryot
c	44	79.2	7.9	2247	8	AAL61177	Aal61177 Actinosyn
c	45	78.8	7.9	9222	8	AAL61171	Aal61171 Actinosyn

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

PS Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC The present invention describes an isolated recombinant polynucleotide
CC (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC synthase (PKS) protein or a fragment comprising one or more domains of
CC the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC host cell (III) comprising (II), or (I) integrated into the cell
CC chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC the native disorazole PKS by the inactivation of one or more domain of
CC the disorazole PKS and/or addition of one or more domain of a non-
CC disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC expression system (VI) capable of producing a disorazole synthase domain
CC in a host cell, where the system comprises an encoding sequence for a
CC disorazole polyketide synthase domain, and the encoding sequence is
CC operably linked to control sequences effective in the cell to produce RNA
CC that is translated into the domain; (7) host cell modified to contain
CC (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC eliminate production of disorazole; (9) an isolated polypeptide encoded
CC by (I); (10) hybrid polyketide synthase comprising one or more
CC polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC of at least 200 base pairs with a sequence identical or substantially
CC identical to a protein encoding region of the 77294 base pair sequence of
CC SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC (I) is useful for producing disorazole polyketides. (III) is useful for
CC producing a polyketide, which involves growing (III) under conditions,
CC where a polyketide synthesised by a PKS comprising a protein encoded by
CC the polynucleotide molecule is produced in the cell. (I) is useful for
CC producing libraries of PKSs, where the polyketide producing colonies are
CC identified and isolated, and the produced polyketides are used
CC collectively in a panel to represent a library or assessed individually
CC for activity. The disorazole polyketide produced by (I) is useful in
CC molecular biology, chemistry, recombinant DNA technology, human and
CC veterinary medicine and agricultural applications. The disorazole
CC polyketide produced by (I) is formulated as a pharmaceutical composition,
CC useful for its activity against microorganisms, in a mammal. The present
CC sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

Query Match 100.0%; Score 1001; DB 12; Length 76994;
Best Local Similarity 100.0%; Pred. No. 2.1e-137;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4000 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 4059

Qy         61 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4060 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 4119

Qy        121 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4120 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 4179

Qy        181 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4180 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 4239

Qy        241 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4240 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 4299

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4300 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 4359

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4360 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 4419

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4420 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 4479

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4480 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 4539

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4540 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 4599

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4600 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 4659

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4660 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 4719

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4720 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 4779

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4780 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 4839

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4840 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 4899

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4900 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 4959

Qy        961 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAGC 1001
              |||||||||||||||||||||||||||||||||||||||||
Db       4960 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAGC 5000
```

RESULT 2
ACL64545
ID ACL64545 standard; DNA; 9867 BP.
XX
AC ACL64545;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus DNA fragment, seq id 1008.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.
XX

```
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 1; SEQ ID NO 1008; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC   set of 1849 contig and singleton sequences comprising coding sequences,
CC   DNA replication elements, promoters and other regulatory elements from
CC   the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO
XX
SQ   Sequence 9867 BP; 1425 A; 3440 C; 3625 G; 1376 T; 0 U; 1 Other;

  Query Match             13.7%;  Score 137.6;  DB 14;  Length 9867;
  Best Local Similarity   54.5%;  Pred. No. 8.2e-12;
  Matches  300;  Conservative    0;  Mismatches  244;  Indels    6;  Gaps    1;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGA 62
              || | |||  |||     |  ||   ||  |||| ||   | | | || ||||||| || |
Db       3497 CCGCCAGCAGGAACAGTGTCACCGCCGTACCGGTGGATTGCCCCGGGTAGCGGAAGGCCA 3556

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              ||| ||| |  ||||||| |||  ||  |||||| |  | ||||||||    |   ||| |
Db       3557 GCACGTCCGCTAGCGACGTGAGCTGCAACTCGCGGGACAGGCGCAGCAGCGGGCGCCACA 3616

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | |  ||  |   || |||||| || ||   ||| ||||| |   | ||| ||  ||  |
Db       3617 GCACGGGTATCAGCAGGCACGCCAGGGTGACGCCCAGGTAGATGCCCAGGTAGCGGAAGC 3676

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              | || ||   |||| |||| |||| |||| |||| | |||||| | |||||| |||| |
Db       3677 CATGCCGGGCCGCGTAGCCCACGCTGCCGAAGTAGGACCACGATGTGGCGTACACGCCCA 3736

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              | |  || |||   |||||||   | | | ||         |||     | | |||| ||
Db       3737 GGGCCAGCGCGTAGACGAGCGGATGCTGGGTGATGCGCGCGGGGATGGCGCCTCGCTCCG 3796

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              | |||   || | |  | ||||||   ||  | || || |  |||   | |||    |
Db       3797 CCGCGTAGGCCACCAGGAAGAGCACCAGCAGGTAGGCGACGGACGCGGCGACGAGCGAAC 3856

Qy        363 CCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCG 422
              | |  |||| |     ||       |     | |||   ||  | |  | | |   | |
Db       3857 CGAGCTCGAGCGTCATGG------GCCACGACTCCGGTGCGCCGCCCAGCCGCCCAGGAG 3910

Qy        423 CGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGC 482
              |  ||  |     | |||||   ||  |  || |  ||||    | |||  |  |
Db       3911 GATGACGCCCAGCCACACCGCGAAGGGCAGGTACCCCGGCGCGCCCTCCGCCAGCCACAG 3970

Qy        483 GAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAG 542
                |   | | |  | ||  | | || |||      ||| || || |     | |||
Db       3971 TCGGCGCAGGGGAGAGCCGAAGAGGAGCACCGCGACGGCGAAGACGAGCAGCGACGGCAT 4030

Qy        543 CACGAGCACG 552
              ||||  |  ||
Db       4031 CACGGGGGCG 4040


RESULT 3
ACL73103/c
ID   ACL73103 standard; DNA; 2973 BP.
XX
AC   ACL73103;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus gene sequence, seq id 9566.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression;
KW   gene; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 9566; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 2973 BP; 357 A; 1039 C; 1095 G; 482 T; 0 U; 0 Other;

  Query Match             13.6%;  Score 135.8;  DB 14;  Length 2973;
  Best Local Similarity   60.4%;  Pred. No. 1.9e-11;
  Matches  224;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGA 62
              || | |||  |||     |  ||   ||  |||| ||   | | | || ||||||| || |
Db        376 CCGCCAGCAGGAACAGTGTCACCGCCGTACCGGTGGATTGCCCCGGGTAGCGGAAGGCCA 317

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              ||| ||| |  ||||||| |||  ||  |||||| |  | ||||||||    |   ||| |
Db        316 GCACGTCCGCTAGCGACGTGAGCTGCAACTCGCGGGACAGGCGCAGCAGCGGGCGCCACA 257

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | |  ||  |   || |||||| || ||   ||| ||||| |   | ||| ||  ||  |
Db        256 GCACGGGTATCAGCAGGCACGCCAGGGTGACGCCCAGGTAGATGCCCAGGTAGCGGAAGC 197

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              | || ||   |||| |||| |||| |||| |||| | |||||| | |||||| |||| |
Db        196 CATGCCGGGCCGCGTAGCCCACGCTGCCGAAGTAGGACCACGATGTGGCGTACACGCCCA 137

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              | |  || |||   |||||||   | | | ||         |||     | | |||| ||
Db        136 GGGCCAGCGCGTAGACGAGCGGATGCTGGGTGATGCGCGCGGGGATGGCGCCTCGCTCCG 77

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              | |||   || | |  | ||||||   ||  | || || |  |||   | |||    |
Db         76 CCGCGTAGGCCACCAGGAAGAGCACCAGCAGGTAGGCGACGGACGCGGCGACGAGCGAAC 17

Qy        363 CCACGTCGATC 373
```

```
Db         16 CGAGCTCGAGC 6

RESULT 4
ABQ31996/c
ID   ABQ31996 standard; DNA; 1172 BP.
XX
AC   ABQ31996;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18587.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP010074.
XX
PR   01-SEP-2000; 2000DE-01043826.
PR   05-SEP-2000; 2000DE-01044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful for
PT   diagnosis and prognosis, comprises selective hybridization of amplicons
PT   from chemically treated DNA.
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one member,
CC   of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC   degree of hybridisation to both classes is determined from the label on
CC   the amplicon. From the ratio of labels hybridised to the two classes of
CC   oligomers, the degree of methylation is calculated. The method is used:
CC   (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC   and of a wide range of diseases, e.g. cancer, disorders of the central
CC   nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC   particularly by detecting mutations or single nucleotide polymorphisms
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 1172 BP; 0 A; 583 C; 586 G; 3 T; 0 U; 0 Other;

  Query Match             12.3%;  Score 123.4;  DB 6;  Length 1172;
  Best Local Similarity   48.5%;  Pred. No. 1.4e-09;
  Matches  482;  Conservative    0;  Mismatches  506;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
Db       1090 CACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1031

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
Db       1030 CGCGCGCGCGCGCGC-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 972

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
Db        971 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 912

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
Db        911 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 852

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
Db        851 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 792

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
Db        791 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 732

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
Db        731 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 672

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
Db        671 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 612

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
Db        611 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 553

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
Db        552 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
Db        492 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
Db        432 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
Db        372 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 313

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
Db        312 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 253

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
Db        252 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 193

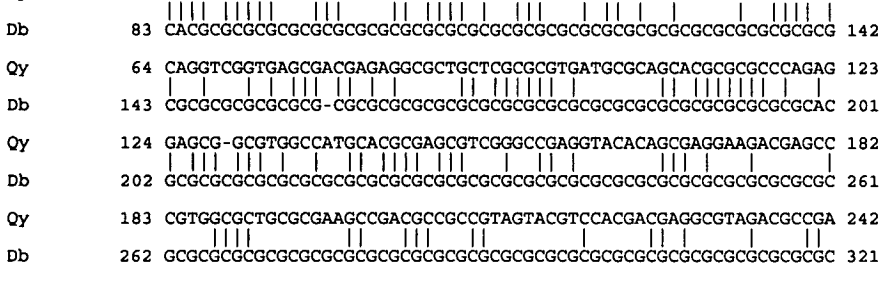
Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
Db        192 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 133

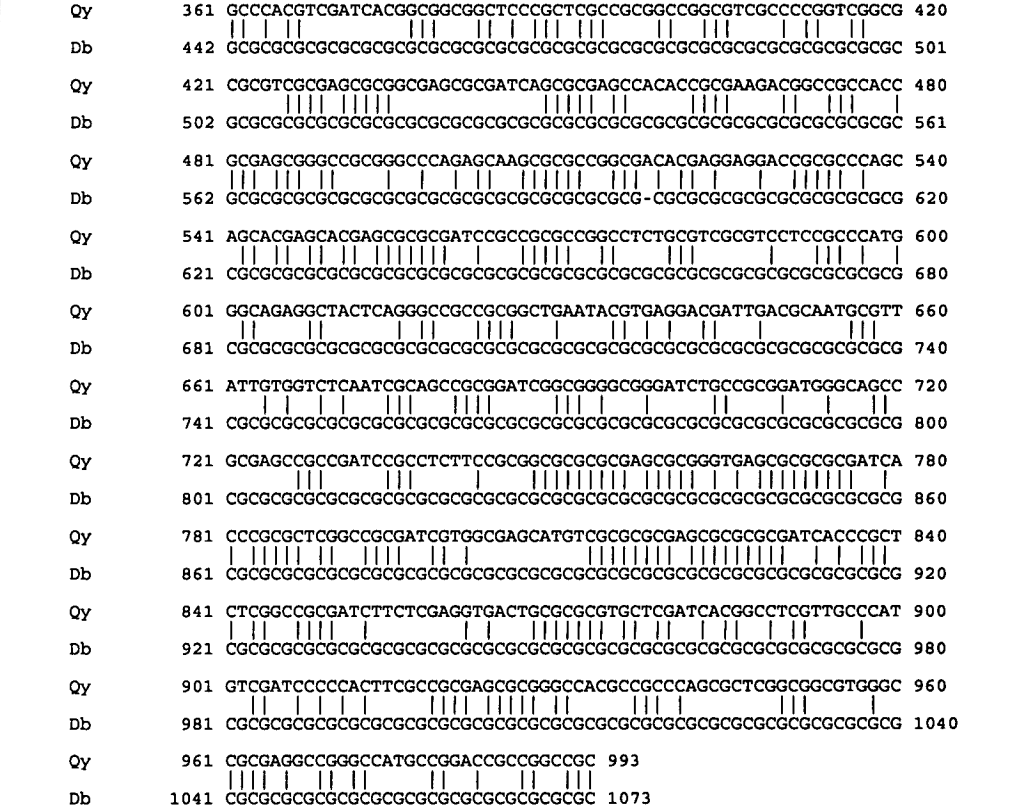
Qy        961 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGC 993
Db        132 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 100

RESULT 5
ABQ31997
ID   ABQ31997 standard; DNA; 1172 BP.
XX
AC   ABQ31997;
XX
DT   12-JUL-2002  (first entry)
XX
```

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18588.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1172 BP; 3 A; 586 C; 583 G; 0 T; 0 U; 0 Other;

Query Match 12.3%; Score 123.4; DB 6; Length 1172;
Best Local Similarity 48.5%; Pred. No. 1.4e-09;
Matches 482; Conservative 0; Mismatches 506; Indels 5; Gaps 5;





RESULT 6
ABQ31995/c
ID ABQ31995 standard; DNA; 1172 BP.
XX
AC ABQ31995;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18586.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.

XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 U; 0 Other;

Query Match 12.2%; Score 122.4; DB 6; Length 1172;
Best Local Similarity 48.5%; Pred. No. 2e-09;
Matches 481; Conservative 0; Mismatches 506; Indels 5; Gaps 5;

```
Qy          6 CGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGAGCA 65
              || |||||   |||    || |||| |  |||   | || |  | |   |  |||| ||
Db       1172 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCG 1113

Qy         66 GGTCGGTGAGCG-ACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
               |   | | |||  || | |     || |||||| |        || ||||||| |    |
Db       1112 CGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053

Qy        125 AGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCC 183
              ||| ||| |  |  || |||| |||   |  || |     ||| |  |     |
Db       1052 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993

Qy        184 GTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAG 243
                ||||        ||  ||||   ||         |     || |     |    || |
Db        992 CGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933

Qy        244 CGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTGC 301
              || | | ||| || || ||| || | |||||| |||||||| |||| || ||||||
Db        932 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873

Qy        302 GCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACG 361
                ||| |||||  ||||          || ||     | |    |||  |  | |  ||
Db        872 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813

Qy        362 CCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC 421
              | | ||          |||   || | ||| |||    ||  |||    | ||  ||
Db        812 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
```

```
Qy        422 GCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCG 481
                ||||  |||||          |||| ||     ||||    ||  |||   ||
Db        752 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693

Qy        482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541
              || ||| ||    |  |  | ||  ||||||  ||| | || |   |  |||||  |
Db        692 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGC 634

Qy        542 GCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGG 601
              || || || || |||||||  |   |||||  ||     |||       |   ||| |  |
Db        633 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 574

Qy        602 GCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTA 661
              ||    ||      | ||   ||||  |    || | |  ||   |        |||
Db        573 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 514

Qy        662 TTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCG 721
              | |  | |   |||   ||||     ||| |   |      ||   |   |   ||
Db        513 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 454

Qy        722 CGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCAC 781
                |||      |||      |    ||||||||| ||||| | | |||||||||  | |
Db        453 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 394

Qy        782 CCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTC 841
              |||||  ||  ||||  || |              |||||||| |||||||||  | | ||| |
Db        393 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 334

Qy        842 TCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATG 901
              ||  ||||  |          | |   |||||||  || ||  | ||  | ||    |
Db        333 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 274

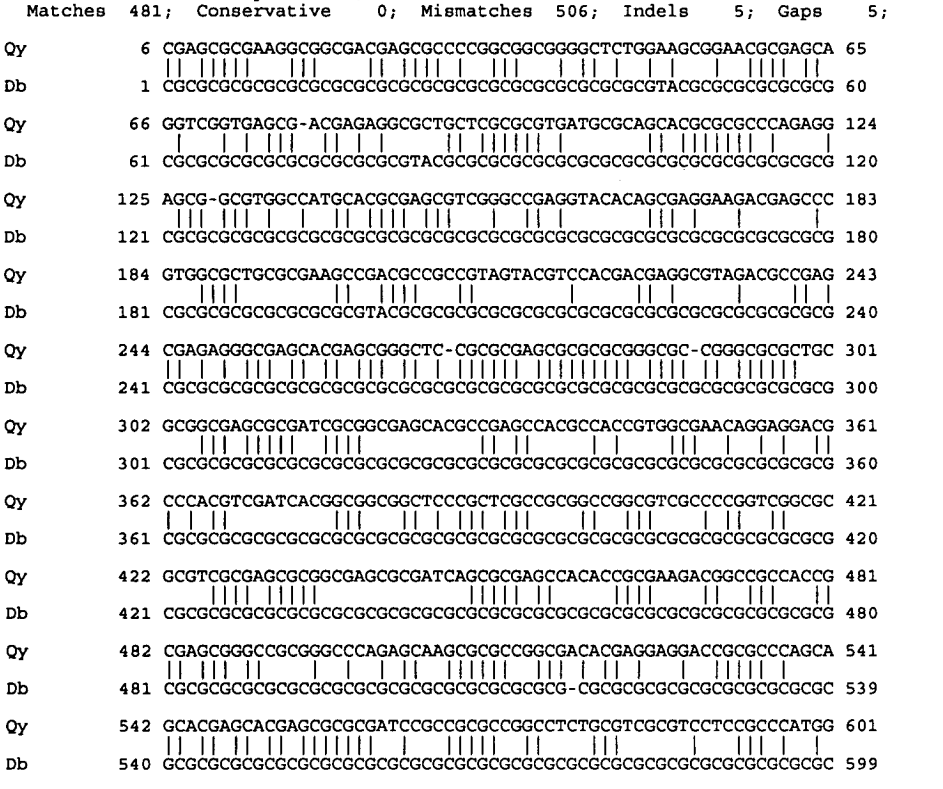
Qy        902 TCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCC 961
              ||  | | | |     |||| |||||  ||      |||  |         |||      | |
Db        273 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 214

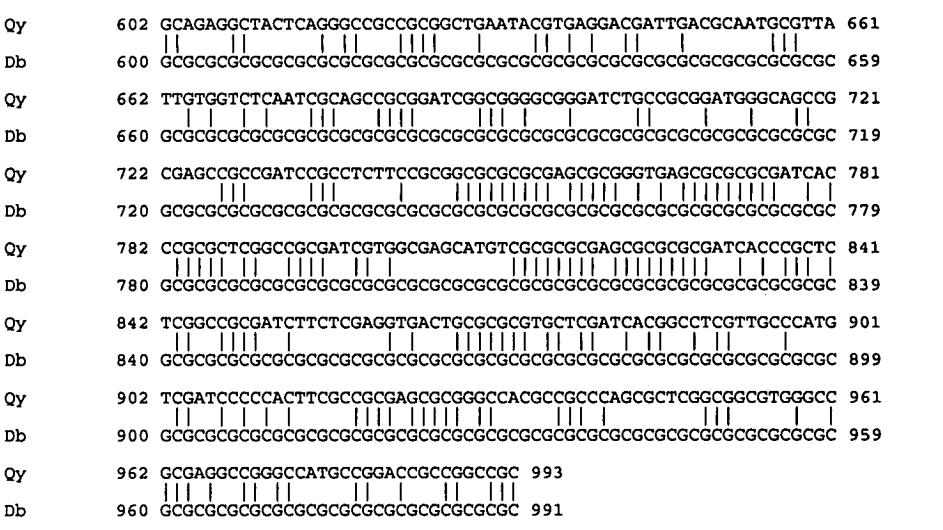
Qy        962 GCGAGGCCGGGCCATGCCGGACCGCCGGCCGC 993
              ||| |  || ||     ||  |   ||  |||
Db        213 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 182
```

RESULT 7
ABQ31994
ID ABQ31994 standard; DNA; 1172 BP.
XX
AC ABQ31994;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18585.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX

PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 U; 0 Other;

Query Match 12.2%; Score 122.4; DB 6; Length 1172;
Best Local Similarity 48.5%; Pred. No. 2e-09;
Matches 481; Conservative 0; Mismatches 506; Indels 5; Gaps 5;





RESULT 8
ABQ31995
ID ABQ31995 standard; DNA; 1172 BP.
XX
AC ABQ31995;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18586.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,

```
CC   of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC   degree of hybridisation to both classes is determined from the label on
CC   the amplicon. From the ratio of labels hybridised to the two classes of
CC   oligomers, the degree of methylation is calculated. The method is used:
CC   (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC   and of a wide range of diseases, e.g. cancer, disorders of the central
CC   nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC   particularly by detecting mutations or single nucleotide polymorphisms
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 U; 0 Other;

  Query Match             12.2%;  Score 122.2;  DB 6;  Length 1172;
  Best Local Similarity   49.2%;  Pred. No. 2.1e-09;
  Matches  458;  Conservative    0;  Mismatches  468;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
              | || |||||   |||    || |||| |  |||   | || |  |     |  |||| |
Db        237 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |   | | ||| || | |      || |||||| |        || |||||||| |
Db        297 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 355

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |        ||| |   |      |
Db        356 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 415

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                ||||        ||   |||   ||        |     || |     |    ||
Db        416 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 475

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| |||||||||| |||| || ||||||
Db        476 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 535

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
                ||| |||||  ||||          || ||     | |    |||  |  | |  |
Db        536 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 595

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||          |||   || | ||| |||    ||  |||    | ||  ||
Db        596 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 655

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
                ||||  |||||              |||| ||     ||||    ||  |||   |
Db        656 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 715

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||    |  |  | ||  ||||||  ||| | || |   |  ||||| |
Db        716 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 774

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              || || || || |||||||  |   |||||  ||     |||       |   ||| |  |
Db        775 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||    ||      | ||   ||||    |    || | |  ||   |         |||
Db        835 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              | |  | |  |||   ||||     ||| |   |     ||   |   |    ||
Db        895 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
                  |||      ||| |   |    ||||||||| ||||| | | |||||||||  |
Db        955 CGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              | ||||| ||  ||||  || |           |||||||| |||||||||  | | |||
Db       1015 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              | ||  ||||  | |      | |   ||||||| || ||  | ||  | |||   |
Db       1075 CGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCG 1134

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGC 931
                ||  | | | |     |||| ||||| ||
Db       1135 CGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1165


RESULT 9
ABQ31996
ID   ABQ31996 standard; DNA; 1172 BP.
XX
AC   ABQ31996;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18587.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP010074.
XX
PR   01-SEP-2000; 2000DE-01043826.
PR   05-SEP-2000; 2000DE-01044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful for
PT   diagnosis and prognosis, comprises selective hybridization of amplicons
PT   from chemically treated DNA.
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one member,
CC   of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC   degree of hybridisation to both classes is determined from the label on
CC   the amplicon. From the ratio of labels hybridised to the two classes of
CC   oligomers, the degree of methylation is calculated. The method is used:
CC   (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC   and of a wide range of diseases, e.g. cancer, disorders of the central
CC   nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC   particularly by detecting mutations or single nucleotide polymorphisms
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 1172 BP; 0 A; 583 C; 586 G; 3 T; 0 U; 0 Other;
```

```
  Query Match             12.2%;  Score 122.2;  DB 6;  Length 1172;
  Best Local Similarity   49.2%;  Pred. No. 2.1e-09;
  Matches  458;  Conservative    0;  Mismatches  468;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
              | || |||||   |||     || |||| |  |||   | || |  |      |  |||| |
Db        237 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |      || |||||| |        || |||||||  |
Db        297 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 355

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |       ||| |   |      |
Db        356 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 415

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                ||||         ||   |||   ||          |     || |   |    ||
Db        416 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 475

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| ||||||||| |||| || ||||||
Db        476 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 535

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
                ||| |||||  ||||           || ||     | |    |||  |  | |  |
Db        536 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 595

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||          |||   || | ||| |||     ||  |||    | ||  ||
Db        596 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 655

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
                ||||  |||||                |||| ||     ||||    ||  |||   |
Db        656 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 715

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||     |  |  | ||  ||||||  ||| | || |    |  |||||  |
Db        716 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 774

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
               || || || || |||||||  |   |||||  ||    |||      |   ||| |  |
Db        775 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||    ||       | ||   ||||    |    || | |  ||    |        |||
Db        835 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
                | |  | |   |||   ||||      ||| |   |      ||    |   |   ||
Db        895 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
                  |||     ||| |   |    ||||||||| ||||| | | |||||||||  |
Db        955 CGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              | ||||| ||  ||||  || |           |||||||| |||||||||  | | |||
Db       1015 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              | ||  |||| | |     | |   |||||| || ||  | ||  | |||   |
Db       1075 CGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCG 1134

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGC 931
                ||  | | | |     |||| |||| ||
Db       1135 CGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1165


RESULT 10
ABQ31997/c
ID   ABQ31997 standard; DNA; 1172 BP.
XX
AC   ABQ31997;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 18588.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP010074.
XX
PR   01-SEP-2000; 2000DE-01043826.
PR   05-SEP-2000; 2000DE-01044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful for
PT   diagnosis and prognosis, comprises selective hybridization of amplicons
PT   from chemically treated DNA.
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one member,
CC   of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC   degree of hybridisation to both classes is determined from the label on
CC   the amplicon. From the ratio of labels hybridised to the two classes of
CC   oligomers, the degree of methylation is calculated. The method is used:
CC   (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC   and of a wide range of diseases, e.g. cancer, disorders of the central
CC   nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC   particularly by detecting mutations or single nucleotide polymorphisms
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 1172 BP; 3 A; 586 C; 583 G; 0 T; 0 U; 0 Other;

  Query Match             12.2%;  Score 122.2;  DB 6;  Length 1172;
  Best Local Similarity   49.2%;  Pred. No. 2.1e-09;
  Matches  458;  Conservative    0;  Mismatches  468;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
              | || |||||   |||     || |||| |  |||   | || |  |      |  |||| |
Db        936 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |      || |||||| |        || |||||||  |
Db        876 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 818

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |       ||| |   |      |
Db        817 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 758
```

```
Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              ||||         ||   |||   ||         |     || |     |    ||
Db        757 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 698

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| ||||||||| |||| || ||||||
Db        697 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 638

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
              ||| |||||  ||||         || ||     | |     |||  |  | |  |
Db        637 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 578

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||          |||   || | ||| |||    ||  |||    | ||  ||
Db        577 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 518

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              ||||  |||||                |||||  ||      ||||     ||  |||   |
Db        517 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 458

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||    |  |  | ||  |||||  ||| | || |    |  |||||
Db        457 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 399

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              || || || || |||||||  |   |||||  ||     |||        |   ||| |  |
Db        398 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 339

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||    ||      | ||   ||||   |    || | |  ||   |       |||
Db        338 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              | |  | |   |||   ||||     ||| |   |     ||    |   |   ||
Db        278 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 219

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              |||      ||| |   |    |||||||| ||||| | | |||||||||  |
Db        218 CGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 159

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              | ||||| ||  ||||  || |          ||||||| |||||||||  | | |||
Db        158 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 99

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              | ||  ||||  | |      | |   ||||||| || ||  | ||  | |||   |
Db         98 CGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCG 39

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGC 931
              ||  | | | |     |||| ||||| ||
Db         38 CGCGCGCGCGCGCGCGCGCGCGCGCGCGC 8
```

RESULT 11
ABQ31994/c
ID ABQ31994 standard; DNA; 1172 BP.
XX
AC ABQ31994;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18585.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1172 BP; 3 A; 583 C; 583 G; 3 T; 0 U; 0 Other;

Query Match 12.2%; Score 122.2; DB 6; Length 1172;
Best Local Similarity 49.2%; Pred. No. 2.1e-09;
Matches 458; Conservative 0; Mismatches 468; Indels 5; Gaps 5;

```
Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
              | || |||||   |||    || |||| |  |||   | ||  |  |     |  |||| |
Db        936 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |      || ||||||  |        || ||||||| |
Db        876 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 818

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |       ||| |   |      |
Db        817 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 758

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                  ||||         ||   |||   ||         |     || |     |    ||
Db        757 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 698

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| ||||||||| |||| || ||||||
Db        697 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 638

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
                  ||| |||||  ||||         || ||     | |     |||  |  | |  |
Db        637 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 578

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||          |||   || | ||| |||    ||  |||    | ||  ||
Db        577 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 518
```

```
Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              |||| |||||              |||||  ||     ||||    ||  |||   |
Db        517 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 458

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||     |  |  | ||  ||||||  ||| | || |   |  ||||| |
Db        457 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 399

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              || || || || |||||||  |   |||||  ||    |||       |   ||| |  |
Db        398 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 339

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||    ||       | ||   ||||    |     || | |  ||   |        |||
Db        338 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              | |  | |   |||   ||||     ||| |   |      ||    |   |   ||
Db        278 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 219

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              |||      |||  |   |    |||||||||  |||||  | |  |||||||||   |
Db        218 CGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 159

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              | |||||  ||  ||||  || |           ||||||||  |||||||||   | | |||
Db        158 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 99

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              | ||  ||||  | |     | |   |||||||  || ||  | ||  | |||   |
Db         98 CGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCG 39

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGC 931
              ||  | | | |     ||||  |||||  ||
Db         38 CGCGCGCGCGCGCGCGCGCGCGCGCGCGC 8
```

RESULT 12
ABD12116
ID ABD12116 standard; DNA; 2217 BP.
XX
AC ABD12116;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #10720.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO78545.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 10720; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 2217 BP; 409 A; 659 C; 812 G; 337 T; 0 U; 0 Other;

 Query Match 11.4%; Score 113.8; DB 11; Length 2217;
 Best Local Similarity 57.0%; Pred. No. 3.1e-08;
 Matches 208; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

```
Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  |||    | |   || || |||||   ||  ||  |||| |||||  ||||
Db       1815 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 1874

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              |||||||  |||||  || ||  | |   |||| || |  || ||  |   |        |
Db       1875 AGAGGTCGGCCAGCGAGGACAGCTGGTATGCGCGTGTCACCCGGAGGATCGGATAGAGCA 1934

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| | |  || | | ||    |      || |||||  |  ||||||||  ||   |
Db       1935 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 1994

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   ||||||| ||| ||| |     ||||||||||||  |
Db       1995 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 2054

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |   || ||||||   ||  | |||  ||||||       | ||||||  |
Db       2055 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 2114

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              |      |||||  ||  |||   |  |||  | || ||  ||   |  |  ||   |
Db       2115 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGCGACGCTGATCAGGATCAGCTGGG 2174

Qy        363 CCACG 367
              |||||
Db       2175 TCAGG 2179
```

RESULT 13
ABD11744/c
ID ABD11744 standard; DNA; 2526 BP.
XX
AC ABD11744;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #10348.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX

```
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO78173.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 10348; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 2526 BP; 400 A; 920 C; 755 G; 451 T; 0 U; 0 Other;

  Query Match            11.4%;  Score 113.8;  DB 11;  Length 2526;
  Best Local Similarity   57.0%;  Pred. No. 3e-08;
  Matches  208;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  |||    | |   || || |||||   ||  ||  |||| ||||| ||||
Db        365 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 306

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
               |||||||  ||||| || ||  | |   |||| || |  || || | |       |
Db        305 AGAGGTCGGCCAGCGAGGACAGCTGGTATGCGCGTGTCACCCGGAGGATCGGATAGAGCA 246

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| | |  || | ||    |      || ||||| |  ||||||||  ||   |
Db        245 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 186

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   ||||||| ||| ||| |    ||||||||||| |
Db        185 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 126

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |   || ||||||   ||  | |||  ||||||      | ||||||  |
Db        125 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 66

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              |      |||||  ||  |||    |  |||  | ||  ||   |   |  ||   |
Db         65 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGCGACGCTGATCAGGATCAGCTGGG 6

Qy        363 CCACG 367
              || |
Db          5 TCAGG 1


RESULT 14
ABD11843/c
ID   ABD11843 standard; DNA; 2991 BP.
XX
AC   ABD11843;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #10447.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO78272.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 10447; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 2991 BP; 489 A; 1072 C; 900 G; 530 T; 0 U; 0 Other;

  Query Match            11.4%;  Score 113.8;  DB 11;  Length 2991;
  Best Local Similarity   57.0%;  Pred. No. 2.9e-08;
  Matches  208;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  |||    | |   || || |||||   ||  ||  |||| ||||| ||||
Db        421 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 362

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
               |||||||  ||||| || ||  | |   |||| || |  || || | |       |
Db        361 AGAGGTCGGCCAGCGAGGACAGCTGGTATGCGCGTGTCACCCGGAGGATCGGATAGAGCA 302

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| | |  || | ||    |      || ||||| |  ||||||||  ||   |
Db        301 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 242

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   ||||||| ||| ||| |    ||||||||||| |
Db        241 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 182
```

```
Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |   || ||||||   ||  | |||  ||||||       | |||||||  |
Db        181 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 122

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
                |     |||||  ||  |||   |  |||  | || ||  ||   |  |  ||   |
Db        121 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGCGACGCTGATCAGGATCAGCTGGG 62

Qy        363 CCACG 367
                |||
Db         61 TCAGG 57


RESULT 15
AAX53491/c
ID   AAX53491 standard; DNA; 114955 BP.
XX
AC   AAX53491;
XX
DT   05-JUL-1999  (first entry)
XX
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
XX
OS   Synthetic.
XX
PN   WO9913886-A1.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US019419.
XX
PR   17-SEP-1997;   97US-0059160P.
PR   09-JUN-1998;   98US-00093972.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 1999-229400/19.
XX
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction.
XX
PS   Disclosure; Page 37; 120pp; English.
XX
CC   The specification describes antisense oligonucleotides (AAX52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene initiation
CC   codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'
CC   -end and the juxta-section between coding and non-coding regions and all
CC   segments of RNAs encoding proteins associated with one or more diseases,
CC   conditions or mixtures. The antisense oligonucleotides may be derived
CC   from sequences AAX55272-74. These multiple target oligonucleotides
CC   (specifically AAX55180-271) can be used for the antisense treatment of
CC   diseases and conditions. Typical diseases and conditions are those
CC   associated with impaired respiration and inflammation, including lung
CC   diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,
CC   acute asthma, allergies, asthma, impeded respiration, respiratory
CC   distress syndrome, pain, cystic fibrosis, pulmonary hypertension,
CC   pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary
CC   disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
CC   colon cancer, breast cancer, lung cancer, pancreatic cancer,
CC   hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
CC   well as all types of cancers which may metastasize or have metastasized
CC   to the lungs, including breast and prostate cancer
XX
SQ   Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

  Query Match            10.8%;  Score 107.8;  DB 2;  Length 114955;
  Best Local Similarity  33.2%;  Pred. No. 1.1e-07;
  Matches  328;  Conservative  101;  Mismatches  556;  Indels    3;  Gaps    2;

Qy         11 GCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCG 70
              ||    | ||||| ||        : |   |||     | |   |       :||  | || ||
Db     105325 GCCCGGCCGGCGGCGCNNHNNNSCGGCCCGGCCGGCGGCGCGNNHNNNSCGGCCCGGCCG 105266

Qy         71 GTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGC 130
              | |     |     :     |  ||  | |  ||||   :   :|| || |    |||||
Db     105265 GCGGCGCGCNNHNNNSCGGCCCGGCCGGCGGCGCGCCNNHNNNSCGGCCCGGCCGGCGGC 105206

Qy        131 GTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGC 190
              | | ||       :||  |     | ||  |  | |:         | |  | |||  |
Db     105205 GCGCCCNNHNNNSCGGCCCGGCCGGCGGCGCGCCCVNNHNNNSCGGCCCGGCCGGCGGCG 105146

Qy        191 TGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGG 250
              || |:   :   :    |||     || |     |: |       || |  |   | |
Db     105145 CGCCCVGNNHNNNSCGGCCCGGCCGGCGGCGCGCCCVGGNNHNNNSCGGCCCGGCCGGCG 105086

Qy        251 GCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGC 310
              ||| || |: ||     :  ||  |  ||   | ||||| :|  |  :   : |
Db     105085 GCGCGCCCVGGCNNHNNNSCGGCCCGGCCGGCGGCGCGCCCVGGCCNNHNNNSCGGCCCG 105026

Qy        311 GCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCG 370
              ||   ||  |||  |: |||:  :   :|  ||  | ||     | |  |:  | :|
Db     105025 GCCGGCGGCGCGCCCVGGCCVNNHNNNSCGGCCCGGCCGGCGGCGCGCCCVGGCCVGNNH 104966

Qy        371 ATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGA 430
              ||||   | |  | || |||| :|||:||  :   :||| | |  ||
Db     104965 NNNSCGGCCCGGCCGGCGGCGCGCCCVGGCCVGCNNHNNNSCGGCCCGGCCGGCGGCGCG 104906

Qy        431 GCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGC 490
              | :||| :|||  :   : || :||     :   :: | |:|| |  :   :   :|| |
Db     104905 CCCVGGCCVGCGNNHNNNSGGCCVGCGGNNHNNNSVGGCCVGCGGNNHNNNS--CVGGCC 104848

Qy        491 CGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCA 550
              :|||| :     |: || :|| |       :  : | :   |  :     || |: ||
Db     104847 VGCGGNNHNNNSCCVGGCCVGCGGNNHNNNSCCCVGGCCVGCGGNNHNNNSGCCCVGGCC 104788

Qy        551 CGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGAGGCT 610
              :| |         | ||: |||:||          :|||  ||    |:   |  :   :
Db     104787 VGCGGNNHNNNSCGCCCVGGCCVGCGGNNHNNNSGCGCCCVGGCCVGCGGNNHNNNSCGC 104728

Qy        611 ACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGTCT 670
              | |:|| |:|| |            ||   :| | :  |   :   : |        :| |
Db     104727 GCCCVGGCCVGCGGNNHNNNSGCGCGCCCVGGCCVGCGGNNHNNNSGGCGCGCCCVGGCC 104668

Qy        671 CAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCGCC 730
              :      :   :|||  || | :||| :|     :   :|  | ||  ||: |  :
Db     104667 VGCGGNNHNNNSCGGCGCGCCCVGGCCVGCGGNNHNNNSGCGGCGCGCCCVGGCCVGCGG 104608

Qy        731 GATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCTCG 790
              :   :  | |      |: || :| |     :  |   |||||  |:  | :||
Db     104607 NNHNNNSGGCGGCGCGCCCVGGCCVGCGGNNHNNNSCGGCGGCGCGCCCVGGCCVGCGGN 104548

Qy        791 GCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCGCG 850
              :   :  | |||| ||  :   |:   |  :   :||   |  |     | :||||:
Db     104547 NHNNNSCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNSGCCGGCGGCGCGCCCVGGCCVGC 104488

Qy        851 ATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGATCCCC 910
              :   :| | | ||||||  :  | :|   :   :  |||   | ||  |||:
Db     104487 GGNNHNNNSGGCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNSCGGCCGGCGGCGCGCCCV 104428

Qy        911 CACTT-CGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGAGGCC 969
              |   ||  :   : ||| |  || ||| |  :|  |:|||  :   :| ||
Db     104427 GGCCVGCGGNNHNNNSCCGGCCGGCGGCGCGCCCVGGCCVGCGGNNHNNNSCCCGGCCGG 104368
```

```
Qy        970 GGGCCATGCCGGACCGCCGGCCGCCTCG 997
              |||    ||:| ||: |||      :|
Db     104367 CGGCGCGCCCVGGCCVGCGGNNHNNNSG 104340
```

Search completed: April 27, 2006, 07:59:58
Job time : 485.694 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 206.521 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 cgccacgagcgcgaaggcgg.....accgccggccgcctcgaagc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	137.6	13.7	9867	3	US-09-902-540-1008	Sequence 1008, Ap
c	2	135.8	13.6	2973	3	US-09-902-540-9566	Sequence 9566, Ap
	3	113.8	11.4	2217	3	US-09-252-991A-10720	Sequence 10720, A
c	4	113.8	11.4	2526	3	US-09-252-991A-10348	Sequence 10348, A
c	5	113.8	11.4	2991	3	US-09-252-991A-10447	Sequence 10447, A
c	6	83.4	8.3	20235	2	US-07-642-734C-3	Sequence 3, Appli
c	7	83.4	8.3	20235	3	US-08-439-009A-3	Sequence 3, Appli
	8	83	8.3	12425	3	US-09-616-289-50	Sequence 50, Appl
	9	83	8.3	12425	3	US-09-976-740-50	Sequence 50, Appl
	10	81.4	8.1	1614	3	US-09-616-289-45	Sequence 45, Appl
	11	81.4	8.1	1614	3	US-09-976-740-45	Sequence 45, Appl
c	12	77.2	7.7	1926	3	US-09-249-585A-4	Sequence 4, Appli
c	13	77.2	7.7	1931	2	US-09-130-114-2	Sequence 2, Appli
c	14	74	7.4	44377	2	US-08-804-227C-7	Sequence 7, Appli
c	15	74	7.4	44377	2	US-08-804-198-1	Sequence 1, Appli
c	16	72.6	7.3	71989	3	US-09-443-501A-2	Sequence 2, Appli
c	17	72.6	7.3	71989	3	US-09-724-889A-2	Sequence 2, Appli
c	18	72.6	7.3	71989	3	US-09-724-876-2	Sequence 2, Appli
	19	71.6	7.2	4257	2	US-08-690-473-1	Sequence 1, Appli
	20	71.6	7.2	4257	3	US-09-259-821A-1	Sequence 1, Appli
	21	71.6	7.2	4257	3	US-08-843-659-1	Sequence 1, Appli
	22	71.6	7.2	4257	3	US-09-825-288A-1	Sequence 1, Appli
c	23	71.6	7.2	12001	2	US-08-458-568A-11	Sequence 11, Appl
c	24	71	7.1	3687	3	US-09-252-991A-1193	Sequence 1193, Ap
c	25	71	7.1	3957	3	US-10-237-551-193	Sequence 193, App
c	26	71	7.1	4266	3	US-09-252-991A-1234	Sequence 1234, Ap
c	27	71	7.1	5877	3	US-10-152-886-54	Sequence 54, Appl
	28	71	7.1	154746	3	US-09-827-688-8	Sequence 8, Appli
c	29	71	7.1	154746	3	US-09-827-688-8	Sequence 8, Appli
c	30	70.6	7.1	801	3	US-09-298-568-3	Sequence 3, Appli
c	31	70.6	7.1	801	3	US-09-894-273-3	Sequence 3, Appli
c	32	70.4	7.0	1288	2	US-08-440-856A-9	Sequence 9, Appli
c	33	70.2	7.0	885	3	US-09-902-540-6856	Sequence 6856, Ap
	34	70.2	7.0	2561	3	US-09-616-289-48	Sequence 48, Appl
	35	70.2	7.0	2561	3	US-09-976-740-48	Sequence 48, Appl
c	36	70.2	7.0	4148	3	US-09-902-540-578	Sequence 578, App
c	37	69.6	7.0	2560	3	US-09-902-540-7617	Sequence 7617, Ap
c	38	69.6	7.0	5017	3	US-09-902-540-748	Sequence 748, App
c	39	69.4	6.9	28958	2	US-08-258-261B-6	Sequence 6, Appli
c	40	69.4	6.9	28958	2	US-08-456-837-6	Sequence 6, Appli
c	41	69.4	6.9	28958	2	US-08-457-342-6	Sequence 6, Appli
c	42	69.4	6.9	28958	2	US-08-457-646A-6	Sequence 6, Appli
c	43	69.4	6.9	28958	2	US-08-458-076A-6	Sequence 6, Appli
c	44	69.4	6.9	28958	2	US-08-764-233A-4	Sequence 4, Appli
c	45	69.4	6.9	28958	2	US-08-457-335A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-902-540-1008
; Sequence 1008, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1008
; LENGTH: 9867
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(9867)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1008

Query Match 13.7%; Score 137.6; DB 3; Length 9867;
Best Local Similarity 54.5%; Pred. No. 6.9e-15;
Matches 300; Conservative 0; Mismatches 244; Indels 6; Gaps 1;

```
Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              || | |||  |||    |  ||   ||  |||| ||   | | | || ||||||| || |
Db       3497 CCGCCAGCAGGAACAGTGTCACCGCCGTACCGGTGGATTGCCCCGGGTAGCGGAAGGCCA 3556

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              ||| ||| |  ||||||| |||  ||  ||||||  |  | |||||||    |   |||| |
Db       3557 GCACGTCCGCTAGCGACGTGAGCTGCAACTCGCGGGACAGGCGCAGCAGCGGGCGCCACA 3616

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | |  ||  |   || |||||| || ||   ||| |||||  |   | |||  ||   ||  |
Db       3617 GCACGGGTATCAGCAGGCACGCCAGGGTGACGCCCAGGTAGATGCCCAGGTAGCGGAAGC 3676

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              | || ||   |||| |||| |||| |||| |||| | |||||| | |||||| |||||| |
Db       3677 CATGCCGGGCCGCGTAGCCCACGCTGCCGAAGTAGGACCACGATGTGGCGTACACGCCCA 3736

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
```

```
Db       3737 GGGCCAGCGCGTAGACGAGCGGATGCTGGGTGATGCGCGCGGGGATGGCGCCTCGCTCCG 3796

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362

Db       3797 CCGCGTAGGCCACCAGGAAGAGCACCAGCAGGTAGGCGACGGACGCGGCGACGAGCGAAC 3856

Qy        363 CCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCG 422

Db       3857 CGAGCTCGAGCGTCATGG------GCCACGACTCCGGTGCGCCGCCCAGCCGCCCAGGAG 3910

Qy        423 CGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGC 482

Db       3911 GATGACGCCCAGCCACACCGCGAAGGGCAGGTACCCCGGCGCGCCCTCCGCCAGCCACAG 3970

Qy        483 GAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAG 542

Db       3971 TCGGCGCAGGGGAGAGCCGAAGAGGAGCACCGCGACGGCGAAGACGAGCAGCGACGGCAT 4030

Qy        543 CACGAGCACG 552

Db       4031 CACGGGGGCG 4040


RESULT 2
US-09-902-540-9566/c
; Sequence 9566, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9566
;   LENGTH: 2973
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-9566

  Query Match            13.6%;  Score 135.8;  DB 3;  Length 2973;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-14;
  Matches  224;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62

Db        376 CCGCCAGCAGGAACAGTGTCACCGCCGTACCGGTGGATTGCCCCGGGTAGCGGAAGGCCA 317

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122

Db        316 GCACGTCCGCTAGCGACGTGAGCTGCAACTCGCGGGACAGGCGCAGCAGCGGGCGCCACA 257

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182

Db        256 GCACGGGTATCAGCAGGCACGCCAGGGTGACGCCCAGGTAGATGCCCAGGTAGCGGAAGC 197

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242

Db        196 CATGCCGGGCCGCGTAGCCCACGCTGCCGAAGTAGGACCACGATGTGGCGTACACGCCCA 137

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302

Db        136 GGGCCAGCGCGTAGACGAGCGGATGCTGGGTGATGCGCGCGGGGATGGCGCCTCGCTCCG 77

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362

Db         76 CCGCGTAGGCCACCAGGAAGAGCACCAGCAGGTAGGCGACGGACGCGGCGACGAGCGAAC 17

Qy        363 CCACGTCGATC 373

Db         16 CGAGCTCGAGC 6


RESULT 3
US-09-252-991A-10720
; Sequence 10720, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10720
;   LENGTH: 2217
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10720

  Query Match            11.4%;  Score 113.8;  DB 3;  Length 2217;
  Best Local Similarity   57.0%;  Pred. No. 7.3e-11;
  Matches  208;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62

Db       1815 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 1874

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122

Db       1875 AGAGGTCGGCCAGCGAGGACAGCTGGTATGCGCGTGTCACCCGGAGGATCGGATAGAGCA 1934

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182

Db       1935 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 1994

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242

Db       1995 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 2054

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302

Db       2055 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 2114

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362

Db       2115 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGCGACGCTGATCAGGATCAGCTGGG 2174

Qy        363 CCACG 367

Db       2175 TCAGG 2179


RESULT 4
US-09-252-991A-10348/c
; Sequence 10348, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
```

```
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10348
;   LENGTH: 2526
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10348

  Query Match             11.4%;  Score 113.8;  DB 3;  Length 2526;
  Best Local Similarity   57.0%;  Pred. No. 7.2e-11;
  Matches  208;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  |||    | |   || || |||||   ||  ||  |||| ||||| ||||
Db        365 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 306

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              |||||||  ||||| || ||  | |   |||| || |  || || |    |
Db        305 AGAGGTCGGCCAGCGAGGACAGCTGGTATGCGCGTGTCACCCGGAGGATCGGATAGAGCA 246

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| | |  || | | ||     |      || ||||| |  |||||||  ||   |
Db        245 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 186

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   ||||||| ||| ||| |    ||||||||||| |
Db        185 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 126

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |   || |||||   ||  | |||  |||||      | |||||  |
Db        125 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 66

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              |      |||||  ||  |||   |  |||  | || ||  ||   |  |  ||   |
Db         65 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGCGACGCTGATCAGGATCAGCTGGG 6

Qy        363 CCACG 367
              ||| |
Db          5 TCAGG 1


RESULT 5
US-09-252-991A-10447/c
; Sequence 10447, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10447
;   LENGTH: 2991
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10447

  Query Match             11.4%;  Score 113.8;  DB 3;  Length 2991;
  Best Local Similarity   57.0%;  Pred. No. 7.1e-11;
  Matches  208;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | |  |||  |||    | |   || || |||||   ||  ||  |||| ||||| ||||
Db        421 CGATCAGCATGAACAGCGTGGTCAGGGCGCCGGCCCAGGTACTGCGGAAACGGAAGGCGA 362

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              |||||||  ||||| || ||  | |   |||| || |  || || |    |
Db        361 AGAGGTCGGCCAGCGAGGACAGCTGGTATGCGCGTGTCACCCGGAGGATCGGATAGAGCA 302

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | | ||| | |  || | | ||     |      || ||||| |  |||||||  ||   |
Db        301 GCACCGGGGCGAGCAGGAAGGCCCCGGAAATCCCCAGGTAGATGGCGAGGAACCCGTAGC 242

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |||   | |  || | ||| ||   ||||||| ||| ||| |    ||||||||||| |
Db        241 CGTACTGGTAGGCCAGGCCTACCGTGCCGTAGAACGCCCAGGCACTGGCGTAGACGCCCA 182

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |||| |||| |   || |||||   ||  | |||  |||||      | |||||  |
Db        181 GCGACAGGGTGTAGACCAGCGGGTGGCGTACCAGCCGGCGCGGCACGTAGCCGCGCTCGG 122

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              |      |||||  ||  |||   |  |||  | || ||  ||   |  |  ||   |
Db        121 TGATCCAGGCGATGCCGAAGAGGGTGGAGAGGTAGGCGACGCTGATCAGGATCAGCTGGG 62

Qy        363 CCACG 367
              ||| |
Db         61 TCAGG 57


RESULT 6
US-07-642-734C-3/c
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  19..10722
;       OTHER INFORMATION:  /codon_start= 19
;       OTHER INFORMATION:  /function= "gene eryA"
;       OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;       OTHER INFORMATION:  6-deoxyerythronolide B"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19..4470
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  97..1482
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1693..2670
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3406..3921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4171..4428
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..10722
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..5847
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6054..7026
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3

  Query Match             8.3%;  Score 83.4;  DB 2;  Length 20235;
  Best Local Similarity   45.9%;  Pred. No. 7.6e-06;
  Matches  442;  Conservative    0;  Mismatches  506;  Indels   15;  Gaps    4;

Qy         21 CGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTGAGCGACG 80
              || ||  | || || | ||       | | |   ||  || |  ||  |||  ||  |
Db       6907 CGGCGGTCTCCTCGACCGCCATCAGCAGCACCGGGTGCGGGCTCACCTCGACGAAGGTGT 6848

Qy         81 AGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTGGCCATGC 140
               || || |||||||  |     ||||| | |  || |  |||| |  ||   |  || |
Db       6847 CGAAGCCCTGCTCGACCAGACCGCGCACCGCCTGCTCGAAGAGCACTGGTCGGCGCAGGT 6788

Qy        141 ACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGCGCGAAGC 200
                 |  |      |  |  || ||  |||  |  | |||   | | ||    | |   |
Db       6787 TGCGGTACCAGTAGGAGGCGTCCATCGCGGAGGTGTCGATGACCTCGCCGGTCACCGTGG 6728

Qy        201 CGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCACGA 260
              |              |   || |   | || |  |  ||| |||  |   ||||||  |
Db       6727 AGTGCAGCGGCACCGAGCCCCGCACGGCGGTGATGGTGCCCAGCTCGGCCGCGAGCCGGT 6668

Qy        261 GCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGATCGCGG 320
```

```
Db       6667 CCTCGATGCGCGCCACCTCCGGGGAGTGGGAGGCGTAGCGCACGGCGATCCGGCGCGGCT 6608

Qy        321 CGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCACGGCGG 380

Db       6607 TCATCTCCCGCGCCTCGGCCTCGGCGAAGAACTCGTCCAGCTCCGCGGTCGGCCCGGCAA 6548

Qy        381 CGGCTCCCGCTC----GCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGCGCGG 436

Db       6547 CGACGCTGGTGCCAGGCCCGTTCACGGCGGCCACCGCGACACGGCCGTCCCAGCGCGCGA 6488

Qy        437 CGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGCGGG 496

Db       6487 GCACGGTCTCCAGCTCGTCGCGGGACGCGCCGACCGACACCATGCCGCCCTGGTCGTCGA 6428

Qy        497 CCCAGAGCAA------GCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCA 550

Db       6427 GCTCGCGCAACACCTGGCTGCGCAGGGCCACCACCTTCGCGGCGGCTTCCAGCGACAGCG 6368

Qy        551 CGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG--GGCAGAGG 608

Db       6367 CGCCCGCCACGTGCGCCGCGGCGATCTCGCCCTGCGAGTGGCCGACGACGGCGGCCGGGG 6308

Qy        609 CTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGT 668

Db       6307 TCACGCCGTGCGCGCGCCACAGCTCGGCCAGCGAGACCATGATCGAGAAGAGCACCGGCT 6248

Qy        669 CTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCG 728

Db       6247 GGAGCACGTCGACGCGCTCGTGCGGGTCGGGACCGCCGTCGCCGCGGACGACGTCGAGCA 6188

Qy        729 CCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCT 788

Db       6187 ACTTCCAGTCGGT---GTGCGGCGAGAGCGCCTCGGCGCACCGGCTCATCGACTCGGCGA 6131

Qy        789 CGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCG 848

Db       6130 ACACCTCGGAGGATTCCAGCAGATCGCGTGCCATGCCGACCCACTGCGCGCCCTGGCCGG 6071

Qy        849 CGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGATCC 908

Db       6070 GGAAGACCAGGACCGGCTTGCGCGGCGCGGAGGTCACCGGCGCGACGGCGTCGGCCGACG 6011

Qy        909 CCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGAGGC 968

Db       6010 GGCGGCCCTCGGCCAGCGCGTCCAGCTCCGCGCACACGCCCGCGCGGTCGTCGCCGAGCA 5951

Qy        969 CGG 971

Db       5950 CGG 5948


RESULT 7
US-08-439-009A-3/c
; Sequence 3, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4171..4428
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..10722
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..5847
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6054..7026
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 4"
;    FEATURE:
```

```
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-08-439-009A-3

  Query Match             8.3%;  Score 83.4;  DB 3;  Length 20235;
  Best Local Similarity   45.9%;  Pred. No. 7.6e-06;
  Matches  442;  Conservative    0;  Mismatches  506;  Indels   15;  Gaps    4;

Qy         21 CGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTGAGCGACG 80
Db       6907 CGGCGGTCTCCTCGACCGCCATCAGCAGCACCGGGTGCGGGCTCACCTCGACGAAGGTGT 6848

Qy         81 AGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTGGCCATGC 140
Db       6847 CGAAGCCCTGCTCGACCAGACCGCGCACCGCCTGCTCGAAGAGCACTGGTCGGCGCAGGT 6788

Qy        141 ACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGCGCGAAGC 200
Db       6787 TGCGGTACCAGTAGGAGGCGTCCATCGCGGAGGTGTCGATGACCTCGCCGGTCACCGTGG 6728

Qy        201 CGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCACGA 260
Db       6727 AGTGCAGCGGCACCGAGCCCCGCACGGCGGTGATGGTGCCCAGCTCGGCCGCGAGCCGGT 6668

Qy        261 GCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGATCGCGG 320
Db       6667 CCTCGATGCGCGCCACCTCCGGGGAGTGGGAGGCGTAGCGCACGGCGATCCGGCGCGGCT 6608

Qy        321 CGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCACGGCGG 380
Db       6607 TCATCTCCCGCGCCTCGGCCTCGGCGAAGAACTCGTCCAGCTCCGCGGTCGGCCCGGCAA 6548

Qy        381 CGGCTCCCGCTC----GCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGCGCGG 436
Db       6547 CGACGCTGGTGCCAGGCCCGTTCACGGCGGCCACCGCGACACGGCCGTCCCAGCGCGCGA 6488

Qy        437 CGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGCGGG 496
Db       6487 GCACGGTCTCCAGCTCGTCGCGGGACGCGCCGACCGACACCATGCCGCCCTGGTCGTCGA 6428

Qy        497 CCCAGAGCAA------GCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCA 550
Db       6427 GCTCGCGCAACACCTGGCTGCGCAGGGCCACCACCTTCGCGGCGGCTTCCAGCGACAGCG 6368

Qy        551 CGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG--GGCAGAGG 608
Db       6367 CGCCCGCCACGTGCGCCGCGGCGATCTCGCCCTGCGAGTGGCCGACGACGGCGGCCGGGG 6308

Qy        609 CTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGT 668
Db       6307 TCACGCCGTGCGCGCGCCACAGCTCGGCCAGCGAGACCATGATCGAGAAGAGCACCGGCT 6248

Qy        669 CTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCG 728
Db       6247 GGAGCACGTCGACGCGCTCGTGCGGGTCGGGACCGCCGTCGCCGCGGACGACGTCGAGCA 6188

Qy        729 CCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCT 788
Db       6187 ACTTCCAGTCGGT---GTGCGGCGAGAGCGCCTCGGCGCACCGGCTCATCGACTCGGCGA 6131

Qy        789 CGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCG 848
Db       6130 ACACCTCGGAGGATTCCAGCAGATCGCGTGCCATGCCGACCCACTGCGCGCCCTGGCCGG 6071

Qy        849 CGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGATCC 908
Db       6070 GGAAGACCAGGACCGGCTTGCGCGGCGCGGAGGTCACCGGCGCGACGGCGTCGGCCGACG 6011

Qy        909 CCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGAGGC 968
Db       6010 GGCGGCCCTCGGCCAGCGCGTCCAGCTCCGCGCACACGCCCGCGCGGTCGTCGCCGAGCA 5951
```

```
Qy        969 CGG 971
              |||
Db       5950 CGG 5948


RESULT 8
US-09-616-289-50
; Sequence 50, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/09/616,289
;  CURRENT FILING DATE:  2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
;   LENGTH: 12425
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-616-289-50

  Query Match             8.3%;  Score 83;  DB 3;  Length 12425;
  Best Local Similarity   46.2%;  Pred. No. 9.2e-06;
  Matches  383;  Conservative    0;  Mismatches  440;  Indels    6;  Gaps    3;

Qy          6 CGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCA 65
              || | |  | ||||| | |    ||||   |  ||  |      |  |  |  ||| | ||
Db       2824 CGGGAGGCATGGCGGGGCCCCCGGCCCTACCCCCGCCGGAGACGGCGGCGGCCGCCACCA 2883

Qy         66 GGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGA 125
              | |||    || |  |  |  |  | |   |  |      |      | |   || | |  |
Db       2884 CGGCGGCCGCCGCCTCGTCGTCCGCCGCTTCCCCGCACTACCAAGAGTGGATCCTGGACA 2943

Qy        126 GCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGT 185
              |  ||    | ||| | || ||   | || |  | ||   | | | |      |   |
Db       2944 CCATCGACTCGCTGCGCTCGCGCAAGGCGCGGCCGGACCTGGAGCGCATCTGCCGGATGG 3003

Qy        186 GGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCG 245
              |||  |   | |      ||  |||| ||  |   | || ||||    |||  | || |
Db       3004 TGCGGCGGCGGCACGGCCCGGAGCCGGAGCGCACGCGCGCCGAGCTCGAGAAACTGATCC 3063

Qy        246 AGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGC----GGGCGCCGGGCGCGCTGC 301
              ||  |      | |  | || || ||   |  | | ||      |    |    ||||  ||
Db       3064 AGCAGCGCGCCGTGCTCCGGGTCAGCTACAAGGGGAGCATCTCGTACCGCAACGCGGCGC 3123

Qy        302 GCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACG 361
              ||| | |||      ||  |||    | ||  |||   ||  |   |||  |      |  |
Db       3124 GCGTCCAGCCGCCCCGGCGCGGAGCCACCCCGCCGGCCCCGCCGCGCGCCCCCCGCGGGG 3183

Qy        362 CCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC 421
              ||| || ||    || || || | | |   ||||   ||| ||  |||| |   || |
Db       3184 CCCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCCACGCCCGCCCCGCCGCCACCGCCCG 3243

Qy        422 GCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCG 481
                ||    ||| ||   | ||  | | |||  || | | ||      | ||||||| |
Db       3244 CGCCCGTCGCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCGCCGCCACAG 3303

Qy        482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541
              |||  |  | |    ||  |||| ||  |   | || ||||    |||  | || |
Db       3304 CGCCCCCCTCGCCTGGCCCCGCGCAGCCGGGCCCCCGCGCGCAGCGGGCCGCGCCCCTGG 3363

Qy        542 GCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCG-CGTCCTCCGCCCATG 600
              | ||    ||  ||||| | ||||  | |||||    ||| |  || ||  | | |
Db       3364 CCGCGCCGCCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCCGGCCCGCGCC 3423

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              |       | | |   | ||||||    ||       |  | ||        ||  ||
Db       3424 GCGCCCCCCCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCGCCACAGCCGC 3483

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
                  |  | ||     |      || ||  || |   | ||||| |  ||   |
Db       3484 CGGCGCCGCCACAGCAGCAGCGCCGCCGCCGCCGCAGCCACAGCCGCCGCCGGAGGGGG 3543

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              ||| |  | ||  ||| |    | |||  || || ||||  | | ||    || ||| |
Db       3544 GCGCGGTG-CGGGCCGGCGGCGCGGCGGCCCGTGAGCCTGCGGGAAGTCGTGCGCTAC 3602

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGC 829
              | || |  | || |||     ||||  |   |  | ||||  |||| ||
Db       3603 CTCGGGGGCAGCGGCGGCGCCGGCGGTCGCCTAACCCGCGGCCGCGTGC 3651


RESULT 9
US-09-976-740-50
; Sequence 50, Application US/09976740
; Patent No. 6878817
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/09/976,740
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
;   LENGTH: 12425
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-740-50

  Query Match             8.3%;  Score 83;  DB 3;  Length 12425;
  Best Local Similarity   46.2%;  Pred. No. 9.2e-06;
  Matches  383;  Conservative    0;  Mismatches  440;  Indels    6;  Gaps    3;

Qy          6 CGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCA 65
              || | |  | ||||| | |    ||||   |  ||  |      |  |  |  ||| | ||
Db       2824 CGGGAGGCATGGCGGGGCCCCCGGCCCTACCCCCGCCGGAGACGGCGGCGGCCGCCACCA 2883

Qy         66 GGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGA 125
              | ||||    || |  |  |  |  | |   |  |      |      | |   || | |  |
Db       2884 CGGCGGCCGCCGCCTCGTCGTCCGCCGCTTCCCCGCACTACCAAGAGTGGATCCTGGACA 2943

Qy        126 GCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGT 185
              |  ||    | ||| | || ||   | || |  | ||   | | | |      |   |
Db       2944 CCATCGACTCGCTGCGCTCGCGCAAGGCGCGGCCGGACCTGGAGCGCATCTGCCGGATGG 3003

Qy        186 GGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCG 245
              |||  |   | |      ||  |||| ||  |   | || ||||    |||  | || |
```

```
Db       3004 TGCGGCGGCGGCACGGCCCGGAGCCGGAGCGCACGCGCGCCGAGCTCGAGAAACTGATCC 3063

Qy        246 AGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGC----GGGCGCCGGGCGCGCTGC 301

Db       3064 AGCAGCGCGCCGTGCTCCGGGTCAGCTACAAGGGGAGCATCTCGTACCGCAACGCGGCGC 3123

Qy        302 GCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACG 361

Db       3124 GCGTCCAGCCGCCCCGGCGCGGAGCCACCCCGCCGGCCCGCCGCGCGCCCCCCGCGGGG 3183

Qy        362 CCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC 421

Db       3184 CCCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCCACGCCCGCCCCGCCGCCACCGCCCG 3243

Qy        422 GCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCG 481

Db       3244 CGCCCGTCGCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCGCCGCCACAG 3303

Qy        482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541

Db       3304 CGCCCCCCTCGCCTGGCCCCGCGCAGCCGGGCCCCCGCGCGCAGCGGGCCGCGCCCCTGG 3363

Qy        542 GCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCG-CGTCCTCCGCCCATG 600

Db       3364 CCGCGCCGCCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCCGGCCCGCGCC 3423

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660

Db       3424 GCGCCCCCCCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCGCCACAGCCGC 3483

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720

Db       3484 CGGCGCCGCCACAGCAGCAGCCGCCGCCGCCGCAGCCACAGCCGCCGCCGGAGGGGG 3543

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780

Db       3544 GCGCGGTG-CGGGCCGGCGGCGCGGCGCGGCCCGTGAGCCTGCGGGAAGTCGTGCGCTAC 3602

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGC 829

Db       3603 CTCGGGGGCAGCGGCGGCGCCGGCGGTCGCCTAACCCGCGGCCGCGTGC 3651


RESULT 10
US-09-616-289-45
; Sequence 45, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/09/616,289
;  CURRENT FILING DATE:  2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 1614
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1614)
US-09-616-289-45

  Query Match             8.1%;  Score 81.4;  DB 3;  Length 1614;
  Best Local Similarity   46.2%;  Pred. No. 1.9e-05;
  Matches  379;  Conservative    0;  Mismatches  436;  Indels    6;  Gaps    3;

Qy          14 AAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTG 73

Db           1 ATGGCGGGGGCCCCCGGCCCTACCCCCGCCGGAGACGGCGGCGGCCGCCACCACGGCGGCC 60

Qy          74 AGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTG 133

Db          61 GCCGCCTCGTCGTCCGCCGCTTCCCCGCACTACCAAGAGTGGATCCTGGACACCATCGAC 120

Qy         134 GCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGC 193

Db         121 TCGCTGCGCTCGCGCAAGGCGCGGCCGGACCTGGAGCGCATCTGCCGGATGGTGCGGCGG 180

Qy         194 GCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCG 253

Db         181 CGGCACGGCCCGGAGCCGGAGCGCACGCGCGCCGAGCTCGAGAAACTGATCCAGCAGCGC 240

Qy         254 AGCACGAGCGGGCTCCGCGCGAGCGCGCGC----GGGCGCCGGGCGCGCTGCGCGGCGAG 309

Db         241 GCCGTGCTCCGGGTCAGCTACAAGGGGAGCATCTCGTACCGCAACGCGGCGCGCGTCCAG 300

Qy         310 CGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTC 369

Db         301 CCGCCCCGGCGCGGAGCCACCCCGCCGGCCCCGCCGCGCGCCCCCCGCGGGGCCCCCGCC 360

Qy         370 GATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCG 429

Db         361 GCCGCCGCCGCCGCCGCGCCGCCGCCCACGCCCGCCCCGCCGCCACCGCCCGCGCCCGTC 420

Qy         430 AGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGG 489

Db         421 GCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCGCCGCCACAGCGCCCCCC 480

Qy         490 CCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGC 549

Db         481 TCGCCTGGCCCCGCGCAGCCGGGCCCCCGCGCGCAGCGGGCCGCGCCCCTGGCCGCGCCG 540

Qy         550 ACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCG-CGTCCTCCGCCCATGGGCAGAGG 608

Db         541 CCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCCGGCCCGCGCCGCGCCCCC 600

Qy         609 CTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGT 668

Db         601 CCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCGCCACAGCCGCCGGCGCCG 660

Qy         669 CTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCG 728

Db         661 CCACAGCAGCAGCAGCCGCCGCCGCCGCAGCCACAGCCGCCGCCGGAGGGGGGCGCGGTG 720

Qy         729 CCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCT 788

Db         721 -CGGGCCGGCGGCGCGGCGCGGCCCGTGAGCCTGCGGGAAGTCGTGCGCTACCTCGGGGG 779

Qy         789 CGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGC 829

Db         780 CAGCGGCGGCGCCGGCGGTCGCCTAACCCGCGGCCGCGTGC 820


RESULT 11
US-09-976-740-45
; Sequence 45, Application US/09976740
; Patent No. 6878817
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
```

```
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/09/976,740
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 1614
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1614)
US-09-976-740-45

  Query Match             8.1%;  Score 81.4;  DB 3;  Length 1614;
  Best Local Similarity   46.2%;  Pred. No. 1.9e-05;
  Matches  379;  Conservative    0;  Mismatches  436;  Indels    6;  Gaps    3;

Qy         14 AAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTG 73
Db          1 ATGGCGGGGCCCCCGGCCCTACCCCCGCCGGAGACGGCGGCGGCCGCCACCACGGCGGCC 60

Qy         74 AGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTG 133
Db         61 GCCGCCTCGTCGTCCGCCGCTTCCCCGCACTACCAAGAGTGGATCCTGGACACCATCGAC 120

Qy        134 GCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGC 193
Db        121 TCGCTGCGCTCGCGCAAGGCGCGGCCGGACCTGGAGCGCATCTGCCGGATGGTGCGGCGG 180

Qy        194 GCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCG 253
Db        181 CGGCACGGCCCGGAGCCGGAGCGCACGCGCGCCGAGCTCGAGAAACTGATCCAGCAGCGC 240

Qy        254 AGCACGAGCGGGCTCCGCGCGAGCGCGCGC----GGGCGCCGGGCGCGCTGCGCGGCGAG 309
Db        241 GCCGTGCTCCGGGTCAGCTACAAGGGGAGCATCTCGTACCGCAACGCGGCGCGCGTCCAG 300

Qy        310 CGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTC 369
Db        301 CCGCCCCGGCGCGGAGCCACCCCGCCGGCCCCGCCGCGCGCCCCCCGCGGGGCCCCCGCC 360

Qy        370 GATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCG 429
Db        361 GCCGCCGCCGCCGCCGCGCCGCCGCCCACGCCCGCCCCGCCGCCACCGCCCGCGCCCGTC 420

Qy        430 AGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGG 489
Db        421 GCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCGCCGCCACAGCGCCCCCC 480

Qy        490 CCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGC 549
Db        481 TCGCCTGGCCCCGCGCAGCCGGGCCCCCGCGCGCAGCGGGCCGCGCCCCTGGCCGCGCCG 540

Qy        550 ACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCG-CGTCCTCCGCCCATGGGCAGAGG 608
Db        541 CCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCCGGCCCGCGCCGCGCCCCC 600

Qy        609 CTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGT 668
Db        601 CCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCGCCACAGCCGCCGGCGCCG 660

Qy        669 CTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCG 728
Db        661 CCACAGCAGCAGCAGCCGCCGCCGCCGCAGCCACAGCCGCCGCCGGAGGGGGGCGCGGTG 720

Qy        729 CCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCT 788
Db        721 -CGGGCCGGCGGCGCGGCGCGGCCCGTGAGCCTGCGGGAAGTCGTGCGCTACCTCGGGGG 779

Qy        789 CGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGC 829
Db        780 CAGCGGCGGCGCCGGCGGTCGCCTAACCCGCGGCCGCGTGC 820


RESULT 12
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert
;  TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
;  FILE REFERENCE: 0867/0D905
;  CURRENT APPLICATION NUMBER: US/09/249,585A
;  CURRENT FILING DATE:  1999-02-11
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein Barr Virus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1926)
;   OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

  Query Match             7.7%;  Score 77.2;  DB 3;  Length 1926;
  Best Local Similarity   50.0%;  Pred. No. 9.7e-05;
  Matches  276;  Conservative    0;  Mismatches  268;  Indels    8;  Gaps    3;

Qy          5 ACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGC 64
Db        996 AGGAGCTGGGGCCGGAGGTGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGAGGACG 937

Qy         65 AGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
Db        936 AGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGG 877

Qy        125 AGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCG 184
Db        876 ACGGGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGG 817

Qy        185 TGGCGCTGCGCGAAGCCGACGCCGCCGT--AGTACGTCCACGACGAGGCGTA-GACGCCG 241
Db        816 AGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACG 757

Qy        242 AGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGC 301
Db        756 GGGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGG 697

Qy        302 GCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACG 361
Db        696 ACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGAGGACG 637

Qy        362 CCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC 421
Db        636 GGGAGGACGGGGAGGACGGGG-----AGGACGAGGACGGGGAGGAGGACGAGGACGGGGA 582

Qy        422 GCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCG 481
Db        581 GGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGG 522

Qy        482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541
```

```
Db        521 GGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGA 462

Qy        542 GCACGAGCACGA 553
              ||| | ||
Db        461 CGGGGAGGAGGA 450
```

```
RESULT 13
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert A.
;  APPLICANT: Damaj, Bassam B.
;  APPLICANT: Robbins, Alan K.
;  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;  TITLE OF INVENTION: From Multiple Transfected Episomes
;  FILE REFERENCE: 0867/1D903US1
;  CURRENT APPLICATION NUMBER: US/09/130,114
;  CURRENT FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1931
;   TYPE: DNA
;   ORGANISM: EBNA
US-09-130-114-2

  Query Match             7.7%;  Score 77.2;  DB 2;  Length 1931;
  Best Local Similarity   50.0%;  Pred. No. 9.7e-05;
  Matches  276;  Conservative    0;  Mismatches  268;  Indels    8;  Gaps    3;

Qy          5 ACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGAGC 64
              | ||||  |    |  |  |||  |   | ||| | ||   ||| | |||     ||
Db        996 AGGAGCTGGGGCCGGAGGTGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGAGGACG 937

Qy         65 AGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
              ||| ||| |||    | ||||    |  || |   || | | || ||| | |   ||||
Db        936 AGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGG 877

Qy        125 AGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCG 184
              |  ||   | | | |   | |||    |||  ||||   |   || ||| |||| |   |
Db        876 ACGGGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGG 817

Qy        185 TGGCGCTGCGCGAAGCCGACGCCGCCGT--AGTACGTCCACGACGAGGCGTA-GACGCCG 241
              ||    |  ||  |  ||||  |  |   || |||   | |||| || | | || | ||
Db        816 AGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACG 757

Qy        242 AGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGC 301
              |   || || ||| |||||   |     |  ||||  || |  ||    || || |  |
Db        756 GGGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGG 697

Qy        302 GCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACG 361
              ||  ||  | ||   ||| ||| |||  |||      | |     || ||  | |||||||
Db        696 ACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGAGGACG 637

Qy        362 CCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC 421
              | | ||   | | ||| |      |  ||  |  |  |  |  | |  || ||| |
Db        636 GGGAGGACGGGGAGGACGGGG-----AGGACGAGGACGGGGAGGAGGACGAGGACGGGGA 582

Qy        422 GCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCG 481
              |    | |||  ||| |||    ||  ||  || |     || | || || | |     || |
Db        581 GGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGG 522

Qy        482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541
              |||  || || || |  ||| | | |       |  | ||||| ||   ||   || |
Db        521 GGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGA 462

Qy        542 GCACGAGCACGA 553
              ||| | ||
Db        461 CGGGGAGGAGGA 450
```

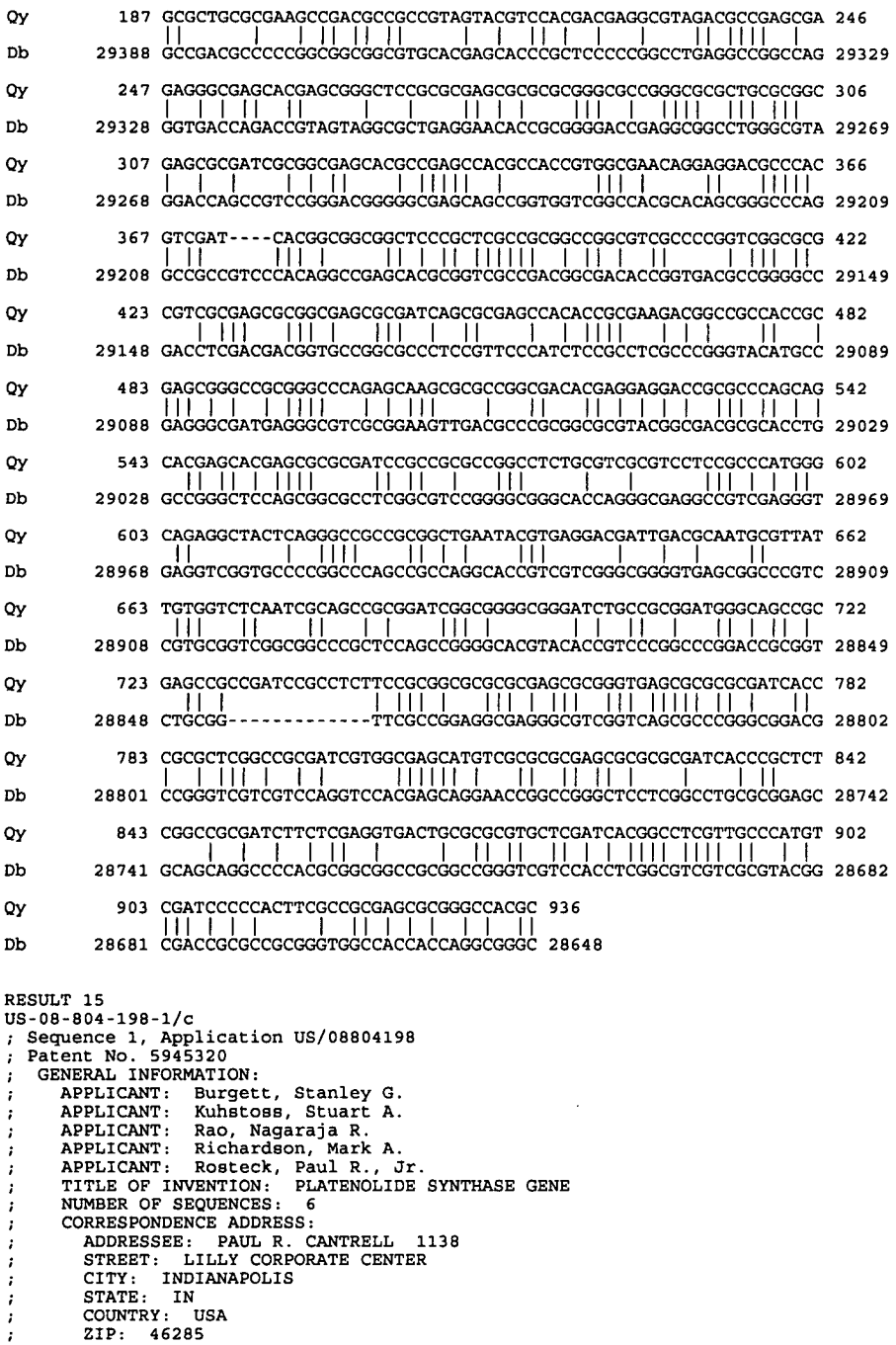
```
RESULT 14
US-08-804-227C-7/c
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             7.4%;  Score 74;  DB 2;  Length 44377;
  Best Local Similarity   45.1%;  Pred. No. 0.00027;
  Matches  421;  Conservative    0;  Mismatches  490;  Indels   23;  Gaps    3;

Qy         13 GAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGT 72
              || | || |  || | | |  | || | |  |    ||||  |  |||||     |  |||
Db      29568 GACGACGTCCGCGCGGGACGTGCCGTCCGCTCCGGTGAAGGCGGTCGCGAAGTCGAGGGT 29509

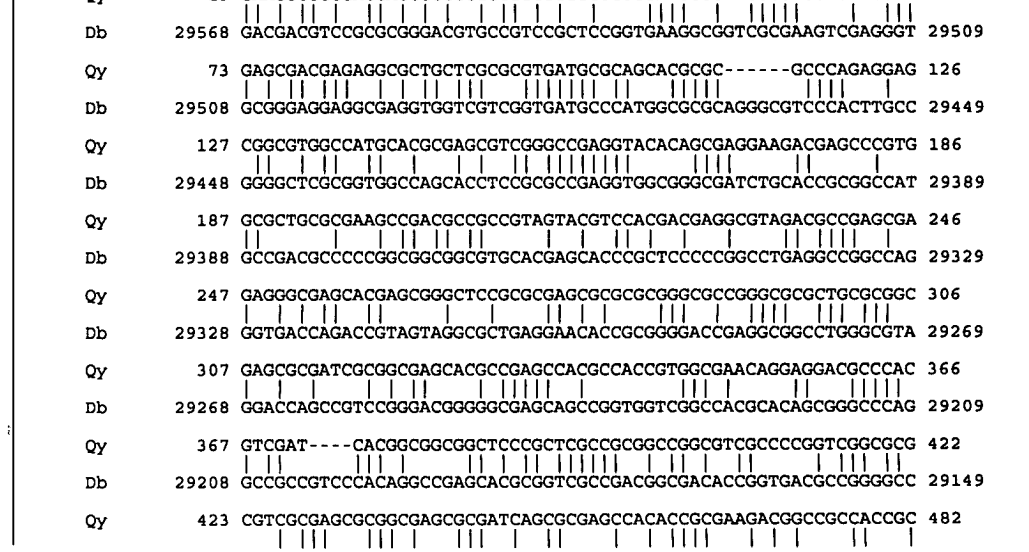
Qy         73 GAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGC------GCCCAGAGGAG 126
              | | || |||  | | ||  |||   ||||||| ||   |||||        ||||   |
Db      29508 GCGGGAGGAGGCGAGGTGGTCGTCGGTGATGCCCATGGCGCGCAGGGCGTCCCACTTGCC 29449

Qy        127 CGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTG 186
              ||  | ||  ||  |   | |  || ||||||||     ||||     ||     |
Db      29448 GGGGCTCGCGGTGGCCAGCACCTCCGCGCCGAGGTGGCGGGCGATCTGCACCGCGGCCAT 29389
```



```
RESULT 15
US-08-804-198-1/c
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             7.4%;  Score 74;  DB 2;  Length 44377;
  Best Local Similarity   45.1%;  Pred. No. 0.00027;
  Matches  421;  Conservative    0;  Mismatches  490;  Indels   23;  Gaps    3;
```



```
Db      29148 GACCTCGACGACGGTGCCGGCGCCCTCCGTTCCCATCTCCGCCTCGCCCGGGTACATGCC 29089

Qy        483 GAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAG 542
              ||| | |  | ||||   | | |||    |   ||   || | | | |  ||| || | |
Db      29088 GAGGGCGATGAGGGCGTCGCGGAAGTTGACGCCCGCGGCGCGTACGGCGACGCGCACCTG 29029

Qy        543 CACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGG 602
              || || | ||||     || ||  |   |||      |   |      ||| | | ||
Db      29028 GCCGGGCTCCAGCGGCGCCTCGGCGTCCGGGGCGGGCACCAGGGCGAGGCCGTCGAGGGT 28969

Qy        603 CAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTAT 662
              ||        |  ||||    || | |    |||       |   |  |    ||
Db      28968 GAGGTCGGTGCCCCGGCCCAGCCGCCAGGCACCGTCGTCGGGCGGGGTGAGCGGCCCGTC 28909

Qy        663 TGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGC 722
              |||   ||    ||   | |    ||| |       | |  ||  |   || | || |
Db      28908 CGTGCGGTCGGCGGCCCGCTCCAGCCGGGGCACGTACACCGTCCCGGCCCGGACCGCGGT 28849

Qy        723 GAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACC 782
              || |               | ||| |   ||| | |||  ||| |||||  || |    ||
Db      28848 CTGCGG------------TTCGCCGGAGGCGAGGGCGTCGGTCAGCGCCCGGGCGGACG 28802

Qy        783 CGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCT 842
              |  | ||| |  | |      ||||||  |   ||  || || |    |    | ||
Db      28801 CCGGGTCGTCGTCCAGGTCCACGAGCAGGAACCGGCCGGGCTCCTCGGCCTGCGCGGAGC 28742

Qy        843 CGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGT 902
                |  |  |  | ||  |     |  || ||  || | | |||| |||| ||  | |
Db      28741 GCAGCAGGCCCCACGCGGCGGCCGCGGCCGGGTCGTCCACCTCGGCGTCGTCGCGTACGG 28682

Qy        903 CGATCCCCCACTTCGCCGCGAGCGCGGGCCACGC 936
              ||| | | |     |  || | | |  | |  ||
Db      28681 CGACCGCGCCGCGGGTGGCCACCACCAGGCGGGC 28648


Search completed: April 27, 2006, 12:42:33
Job time : 207.521 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 1003.57 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 cgccacgagcgcgaaggcgg.....accgccggccgcctcgaagc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1001	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
c	2	123.4	12.3	1172	8	US-10-363-345A-18587	Sequence 18587, A
	3	123.4	12.3	1172	8	US-10-363-345A-18588	Sequence 18588, A
c	4	123.4	12.3	1172	9	US-10-363-483A-18587	Sequence 18587, A
	5	123.4	12.3	1172	9	US-10-363-483A-18588	Sequence 18588, A
	6	122.4	12.2	1172	8	US-10-363-345A-18585	Sequence 18585, A
c	7	122.4	12.2	1172	8	US-10-363-345A-18586	Sequence 18586, A
	8	122.4	12.2	1172	9	US-10-363-483A-18585	Sequence 18585, A
c	9	122.4	12.2	1172	9	US-10-363-483A-18586	Sequence 18586, A
c	10	122.2	12.2	1172	8	US-10-363-345A-18585	Sequence 18585, A
	11	122.2	12.2	1172	8	US-10-363-345A-18586	Sequence 18586, A
	12	122.2	12.2	1172	8	US-10-363-345A-18587	Sequence 18587, A
c	13	122.2	12.2	1172	8	US-10-363-345A-18588	Sequence 18588, A
c	14	122.2	12.2	1172	9	US-10-363-483A-18585	Sequence 18585, A
	15	122.2	12.2	1172	9	US-10-363-483A-18586	Sequence 18586, A
	16	122.2	12.2	1172	9	US-10-363-483A-18587	Sequence 18587, A
c	17	122.2	12.2	1172	9	US-10-363-483A-18588	Sequence 18588, A
c	18	104.6	10.4	1951	8	US-10-363-345A-5847	Sequence 5847, Ap
	19	104.6	10.4	1951	8	US-10-363-345A-5848	Sequence 5848, Ap
c	20	104.6	10.4	1951	9	US-10-363-483A-5847	Sequence 5847, Ap
	21	104.6	10.4	1951	9	US-10-363-483A-5848	Sequence 5848, Ap
	22	102	10.2	1951	8	US-10-363-345A-5845	Sequence 5845, Ap
c	23	102	10.2	1951	8	US-10-363-345A-5846	Sequence 5846, Ap
	24	102	10.2	1951	9	US-10-363-483A-5845	Sequence 5845, Ap
c	25	102	10.2	1951	9	US-10-363-483A-5846	Sequence 5846, Ap
	26	92.8	9.3	3024	9	US-10-939-036-35	Sequence 35, Appl
	27	90.4	9.0	1951	8	US-10-363-345A-5847	Sequence 5847, Ap
c	28	90.4	9.0	1951	8	US-10-363-345A-5848	Sequence 5848, Ap
	29	90.4	9.0	1951	9	US-10-363-483A-5847	Sequence 5847, Ap
c	30	90.4	9.0	1951	9	US-10-363-483A-5848	Sequence 5848, Ap
c	31	87.4	8.7	5223	7	US-10-282-122A-14103	Sequence 14103, A
c	32	87	8.7	88421	3	US-09-976-059-1	Sequence 1, Appli
c	33	86.4	8.6	1951	8	US-10-363-345A-5845	Sequence 5845, Ap
	34	86.4	8.6	1951	8	US-10-363-345A-5846	Sequence 5846, Ap
c	35	86.4	8.6	1951	9	US-10-363-483A-5845	Sequence 5845, Ap
	36	86.4	8.6	1951	9	US-10-363-483A-5846	Sequence 5846, Ap
c	37	84.8	8.5	2118	7	US-10-282-122A-14081	Sequence 14081, A
c	38	83.6	8.4	1041	7	US-10-437-963-8597	Sequence 8597, Ap
c	39	83.4	8.3	1371	7	US-10-282-122A-14254	Sequence 14254, A
	40	83	8.3	12425	3	US-09-976-740-50	Sequence 50, Appl
	41	83	8.3	12425	5	US-10-023-529-50	Sequence 50, Appl
	42	83	8.3	12425	5	US-10-023-523-50	Sequence 50, Appl
	43	83	8.3	12425	7	US-10-616-187-50	Sequence 50, Appl
	44	83	8.3	12425	7	US-10-671-242-50	Sequence 50, Appl
	45	81.4	8.1	1614	3	US-09-976-740-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 1001; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 3.6e-236;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4000 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 4059

Qy         61 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4060 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 4119

Qy        121 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4120 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 4179
```

```
Qy        181 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4180 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 4239

Qy        241 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4240 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 4299

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4300 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 4359

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4360 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 4419

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4420 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 4479

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4480 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 4539

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4540 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 4599

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4600 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 4659

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4660 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 4719

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4720 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 4779

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4780 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 4839

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4840 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 4899

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4900 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 4959

Qy        961 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAGC 1001
              |||||||||||||||||||||||||||||||||||||||||
Db       4960 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAGC 5000


RESULT 2
US-10-363-345A-18587/c
; Sequence 18587, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18587
;  LENGTH: 1172
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;  OTHER INFORMATION: CpG-island No: 18587
US-10-363-345A-18587

  Query Match             12.3%;  Score 123.4;  DB 8;  Length 1172;
  Best Local Similarity   48.5%;  Pred. No. 6.3e-21;
  Matches  482;  Conservative    0;  Mismatches  506;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
              |||| |||||   |||   || |||| |  |||  | || |  |     |  |||| |
Db       1090 CACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1031

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |     || |||||| |       || |||||| |  |
Db       1030 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 972

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| | |  || |||| |||   | || |       ||| |   |      |
Db        971 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 912

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              ||||       ||  |||  ||  | |  |  |  || | |   | |   ||
Db        911 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 852

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| ||||||||| |||| || ||||||
Db        851 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 792

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
              ||| |||||  ||||           || ||      | |    |||  |  | |  |
Db        791 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 732

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||           |||   || | ||| |||    ||  |||    | ||  ||
Db        731 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 672

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              ||||  |||||           ||||  ||     ||||    ||  |||    |
Db        671 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 612

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||    |  |  | ||  ||||||  ||| | || |   |  |||||
Db        611 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 553

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              || || || || |||||||  |   |||||  ||    |||      |   ||| | |
Db        552 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||    ||       | ||   ||||    |   || | |  ||   |      |||
Db        492 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              | |  | |   |||   ||||      |||  |   |      ||    |   |   ||
Db        432 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              |||     |||     |    ||||||||| ||||| | | |||||||||  |
Db        372 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 313

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              | ||||| ||  ||||  || |           |||||||| |||||||||  | | |||
Db        312 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 253

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              | ||  ||||  |       | |   |||||| || ||  | ||  | ||    |
Db        252 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 193

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
```

```
Db        192 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 133

Qy        961 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGC 993
              |||| |  || ||    ||  |   ||  |||
Db        132 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 100


RESULT 3
US-10-363-345A-18588
; Sequence 18588, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18588
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18588
US-10-363-345A-18588

  Query Match             12.3%;  Score 123.4;  DB 8;  Length 1172;
  Best Local Similarity   48.5%;  Pred. No. 6.3e-21;
  Matches  482;  Conservative    0;  Mismatches  506;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGAG 63
              |||| |||||   |||    || |||| |  |||   | || |  |     |  |||| |
Db         83 CACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |     || |||||| |       || ||||||| |  |
Db        143 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 201

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |       ||| |   |      |
Db        202 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 261

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                ||||        ||  |||  ||       |    || |     |    ||
Db        262 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 321

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| |||||||| |||| || ||||||
Db        322 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 381

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
                ||| |||||  ||||        || ||    | |    |||  |  | |  |
Db        382 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 441

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||          |||   || | ||| |||    ||  |||    | ||  ||
Db        442 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 501

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
                |||| ||||                ||||| ||      ||||    ||  |||   |
Db        502 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 561

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||    |  |  | ||  ||||||  ||| | || |    |  ||||| |
Db        562 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 620

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
                || || || || |||||||  |   |||||  ||    |||       |   ||| |  |

Db        621 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
                ||    ||      | ||   ||||   |   || | |  ||   |          |||
Db        681 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
                | |  | |   |||   ||||      ||| |   |     ||    |   |   ||
Db        741 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
                  |||      |||       |     ||||||||| ||||| | | |||||||||  |
Db        801 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
                | ||||| ||  ||||  || |                ||||||||  |||||||||  | | |||
Db        861 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 920

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
                | ||  ||||  |          | |   ||||||| || ||   | ||  | ||    |
Db        921 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 980

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
                ||  | | | |     |||| ||||| ||      ||| |          |||     |
Db        981 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040

Qy        961 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGC 993
              |||| |  || ||    ||  |   ||  |||
Db       1041 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1073


RESULT 4
US-10-363-483A-18587/c
; Sequence 18587, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18587
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18587
US-10-363-483A-18587

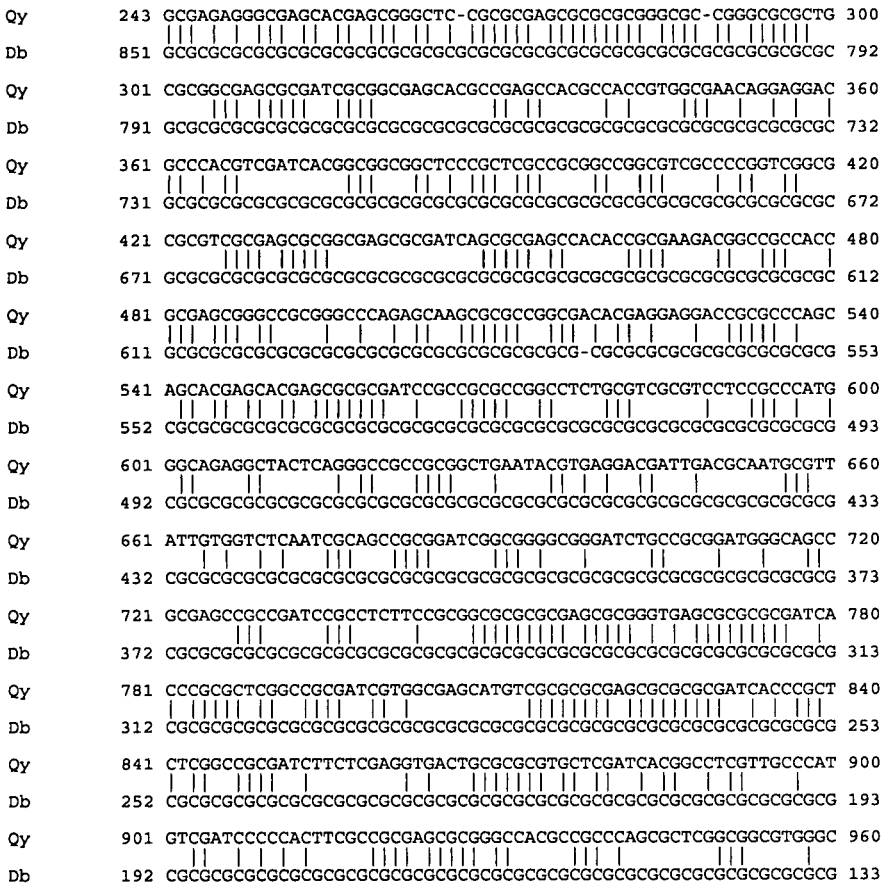
  Query Match             12.3%;  Score 123.4;  DB 9;  Length 1172;
  Best Local Similarity   48.5%;  Pred. No. 6.3e-21;
  Matches  482;  Conservative    0;  Mismatches  506;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGAG 63
              |||| |||||   |||    || |||| |  |||   | || |  |     |  |||| |
Db       1090 CACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1031

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |     || |||||| |       || ||||||| |  |
Db       1030 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 972

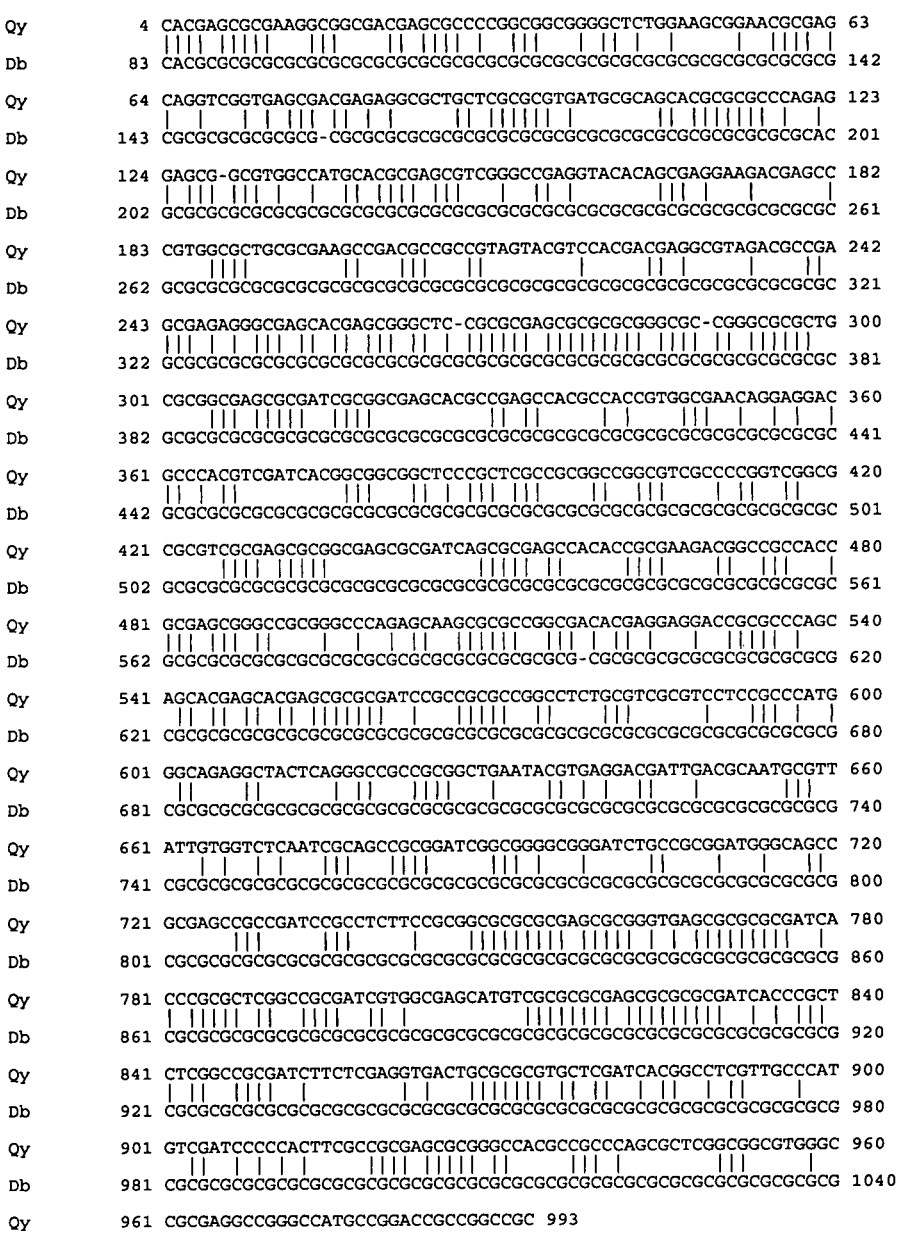
Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |       ||| |   |      |
Db        971 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 912

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                ||||        ||  |||  ||       |    || |     |    ||
Db        911 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 852
```



RESULT 5
US-10-363-483A-18588
; Sequence 18588, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18588
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18588
US-10-363-483A-18588

Query Match 12.3%; Score 123.4; DB 9; Length 1172;
Best Local Similarity 48.5%; Pred. No. 6.3e-21;
Matches 482; Conservative 0; Mismatches 506; Indels 5; Gaps 5;



```
Db       1041 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1073
```

RESULT 6
US-10-363-345A-18585
; Sequence 18585, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18585
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18585
US-10-363-345A-18585

Query Match 12.2%; Score 122.4; DB 8; Length 1172;
Best Local Similarity 48.5%; Pred. No. 1.1e-20;
Matches 481; Conservative 0; Mismatches 506; Indels 5; Gaps 5;

```
Qy          6 CGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCA 65
Db          1 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCG 60

Qy         66 GGTCGGTGAGCG-ACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
Db         61 CGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120

Qy        125 AGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCC 183
Db        121 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180

Qy        184 GTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAG 243
Db        181 CGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240

Qy        244 CGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTGC 301
Db        241 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300

Qy        302 GCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACG 361
Db        301 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360

Qy        362 CCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC 421
Db        361 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420

Qy        422 GCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCG 481
Db        421 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480

Qy        482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541
Db        481 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGC 539

Qy        542 GCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGG 601
Db        540 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 599

Qy        602 GCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTA 661
```

```
Db        600 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 659

Qy        662 TTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCG 721
Db        660 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 719

Qy        722 CGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCAC 781
Db        720 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 779

Qy        782 CCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTC 841
Db        780 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 839

Qy        842 TCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATG 901
Db        840 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 899

Qy        902 TCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCC 961
Db        900 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 959

Qy        962 GCGAGGCCGGGCCATGCCGGACCGCCGGCCGC 993
Db        960 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 991

RESULT 7
US-10-363-345A-18586/c
; Sequence 18586, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18586
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18586
US-10-363-345A-18586

Query Match 12.2%; Score 122.4; DB 8; Length 1172;
Best Local Similarity 48.5%; Pred. No. 1.1e-20;
Matches 481; Conservative 0; Mismatches 506; Indels 5; Gaps 5;

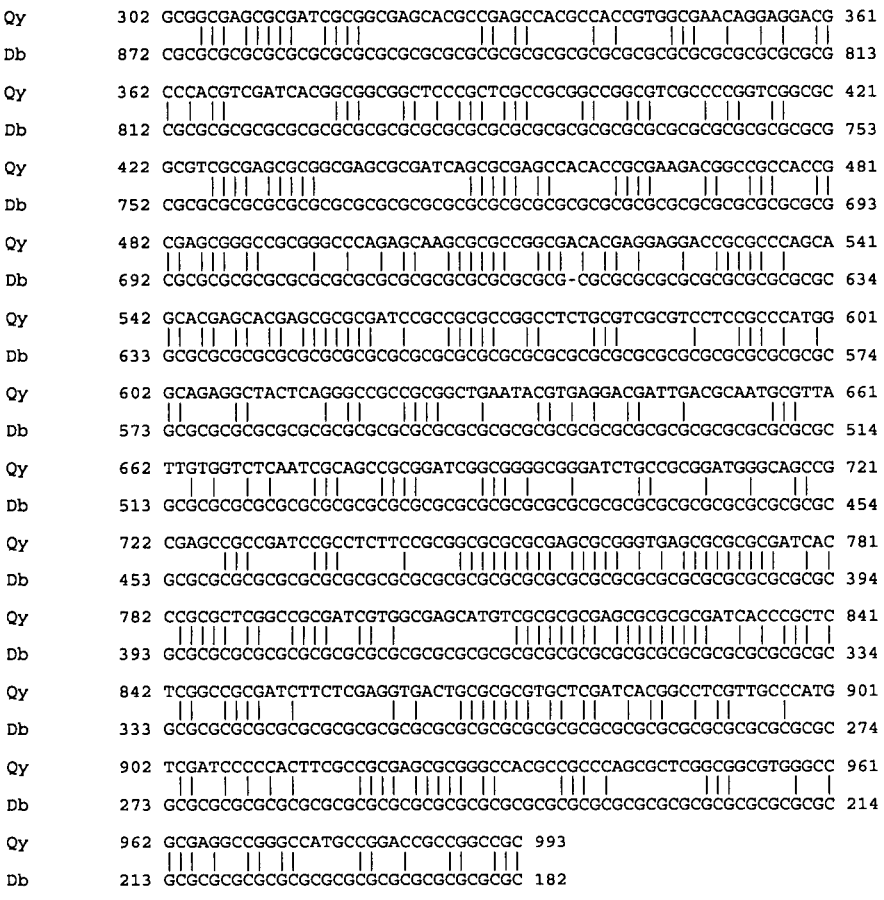
```
Qy 6 CGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCA 65
Db 1172 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCG 1113

Qy 66 GGTCGGTGAGCG-ACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
Db 1112 CGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053

Qy 125 AGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCC 183
Db 1052 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993

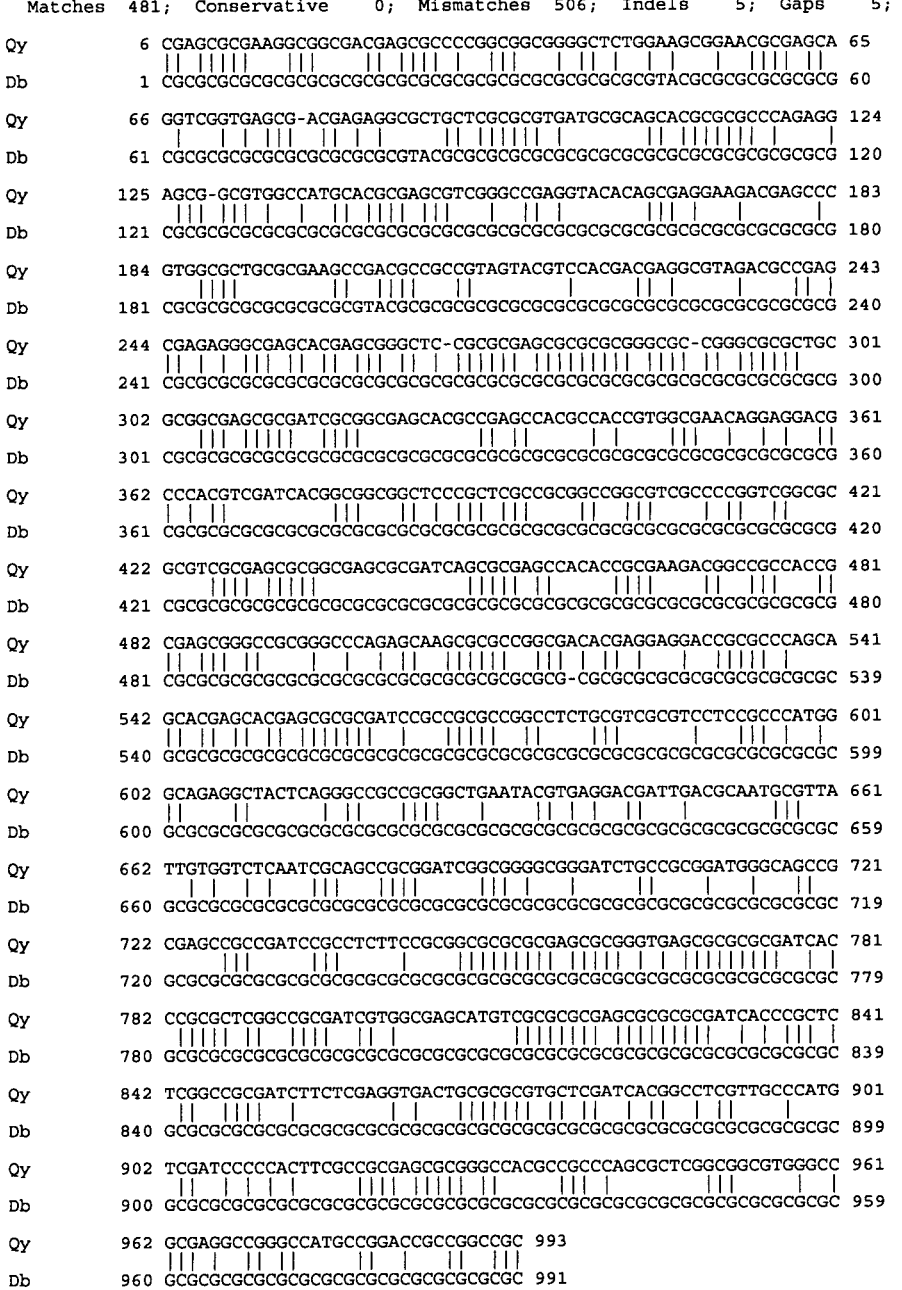
Qy 184 GTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAG 243
Db 992 CGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933

Qy 244 CGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTGC 301
Db 932 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
```
```



RESULT 8
US-10-363-483A-18585
; Sequence 18585, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18585
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18585
US-10-363-483A-18585

Query Match 12.2%; Score 122.4; DB 9; Length 1172;
Best Local Similarity 48.5%; Pred. No. 1.1e-20;
Matches 481; Conservative 0; Mismatches 506; Indels 5; Gaps 5;



RESULT 9
US-10-363-483A-18586/c
; Sequence 18586, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18586
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18586
US-10-363-483A-18586

Query Match 12.2%; Score 122.4; DB 9; Length 1172;
Best Local Similarity 48.5%; Pred. No. 1.1e-20;
Matches 481; Conservative 0; Mismatches 506; Indels 5; Gaps 5;

```
Qy          6 CGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCA 65
              || |||||   |||    || |||| |  |||   | || |  | |   | ||||  ||
Db       1172 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCG 1113

Qy         66 GGTCGGTGAGCG-ACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
                |  | | |||  || | |      || ||||||  |        || |||||||  |    |
Db       1112 CGCGCGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053

Qy         125 AGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCC 183
                ||| ||| |  |  || |||| |||   |  || |      ||| |   |       |
Db       1052 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993

Qy         184 GTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAG 243
                ||||        ||  ||||   ||         |    || |     |    || |
Db       992 CGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933

Qy         244 CGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTGC 301
               || | | ||| || || ||| || | |||||| |||||||| |||| || ||||||
Db       932 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873

Qy         302 GCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACG 361
                ||| |||||  ||||         || ||     | |    |||  |  | |  ||
Db       872 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813

Qy         362 CCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC 421
               | | ||          |||  || | ||| |||     ||  |||    | ||  ||
Db       812 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753

Qy         422 GCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCG 481
                ||||  |||||              |||||  ||     ||||    ||  |||   ||
Db       752 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693

Qy         482 CGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCA 541
               || ||| ||    |  |  | ||  ||||||  ||| | || |   |  |||||  |
Db       692 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGC 634

Qy         542 GCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGG 601
               || || || || |||||||  |   |||||  ||     |||    |  |||  | |
Db       633 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 574

Qy         602 GCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTA 661
               ||   ||    | ||   ||||   |    || | |  ||    |        |||
Db       573 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 514

Qy         662 TTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCG 721
                 | |  | |   |||   ||||     ||| |   |      ||    |   |   ||
```

```
Db       513 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 454

Qy         722 CGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCAC 781
                   |||     |||      |     |||||||| |||||  | |  |||||||||  | |
Db       453 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 394

Qy         782 CCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGCGATCACCCGCTC 841
               |||||  ||  ||||  || |          ||||||||  |||||||||  | |  ||| |
Db       393 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 334

Qy         842 TCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATG 901
               ||  ||||  |          | |   ||||||| ||  ||  | ||  | ||    |
Db       333 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 274

Qy         902 TCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCC 961
               ||  | | | |      ||||  |||||  ||      |||  |        |||      |  |
Db       273 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 214

Qy         962 GCGAGGCCGGGCCATGCCGGACCGCCGGCCGC 993
               |||  |  || ||    ||  |   ||  |||
Db       213 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 182
```

RESULT 10
US-10-363-345A-18585/c
; Sequence 18585, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18585
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18585
US-10-363-345A-18585

Query Match 12.2%; Score 122.2; DB 8; Length 1172;
Best Local Similarity 49.2%; Pred. No. 1.2e-20;
Matches 458; Conservative 0; Mismatches 468; Indels 5; Gaps 5;

```
Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGAG 63
              | || |||||   |||    || |||| |  |||   | || |  |     |  |||| |
Db       936 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877

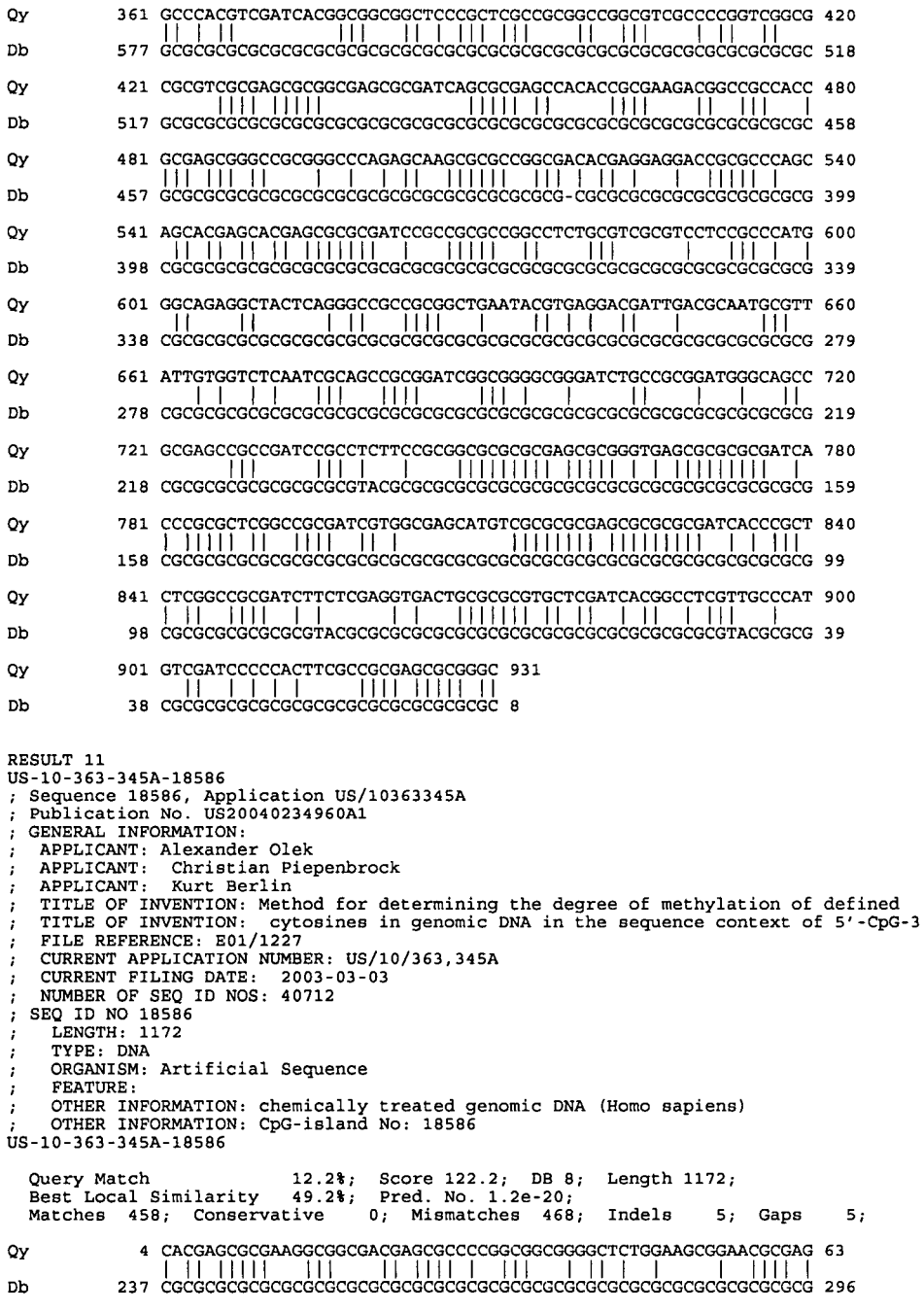
Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
               |  |    | | ||| || | |      || ||||||  |        || |||||||  |
Db       876 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 818

Qy         124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
               | ||| ||| |  |  || |||| |||   |  || |      |||  |   |      |
Db       817 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 758

Qy         183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                  ||||        ||  |||  ||         |    || |     |    ||
Db       757 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 698

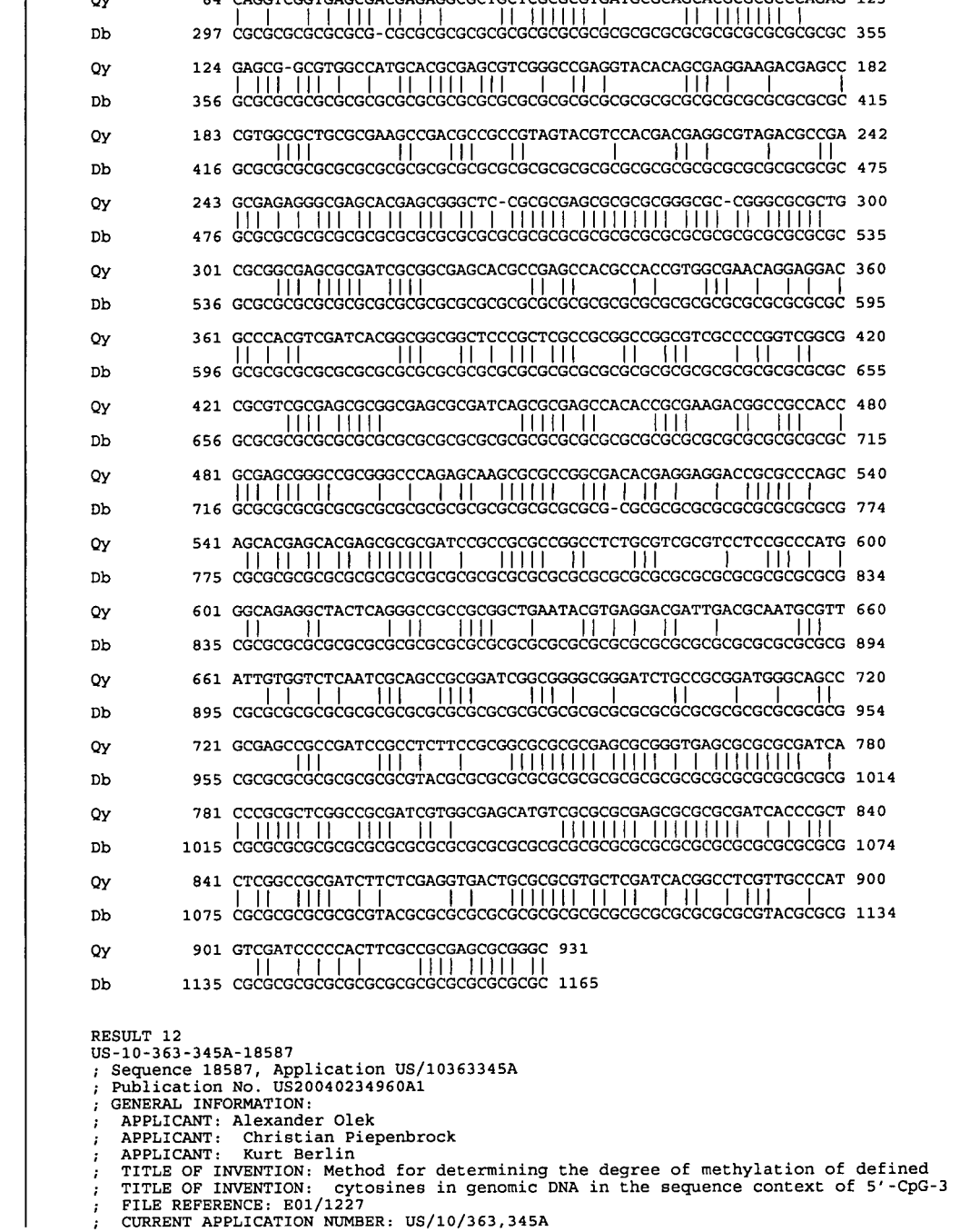
Qy         243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
               ||| | | ||| || || ||| || | |||||| |||||||| |||| || ||||||
Db       697 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 638

Qy         301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
                  ||| |||||  ||||         || ||     | |    |||  |  | |  |
Db       637 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 578
```



RESULT 11
US-10-363-345A-18586
; Sequence 18586, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18586
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18586
US-10-363-345A-18586

Query Match 12.2%; Score 122.2; DB 8; Length 1172;
Best Local Similarity 49.2%; Pred. No. 1.2e-20;
Matches 458; Conservative 0; Mismatches 468; Indels 5; Gaps 5;



RESULT 12
US-10-363-345A-18587
; Sequence 18587, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A

```
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18587
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18587
US-10-363-345A-18587

  Query Match            12.2%;  Score 122.2;  DB 8;  Length 1172;
  Best Local Similarity   49.2%;  Pred. No. 1.2e-20;
  Matches  458;  Conservative    0;  Mismatches  468;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
              | || |||||   |||    || |||| |  |||   | || |  |     |  |||| |
Db        237 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |     || |||||| |        || |||||||
Db        297 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 355

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |       ||| |   |      |
Db        356 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 415

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                ||||        ||  |||  ||         |      || |     |    ||
Db        416 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 475

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| |||||||| |||| || ||||||
Db        476 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 535

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
                ||| |||||  ||||        || ||     | |    |||  |  | |  |
Db        536 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 595

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||          |||   || | ||| |||     ||  |||    | ||  ||
Db        596 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 655

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
                |||| |||||                |||| ||     ||||    ||  |||   |
Db        656 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 715

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||    |  |  | ||  ||||||  ||| | || |   |  |||||
Db        716 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCG 774

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
                || || ||  |||||||  |   |||||  ||    |||       |   ||| |  |
Db        775 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
                ||    ||       | ||   ||||   |    || | |  ||   |       |||
Db        835 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
                | |  | |   |||   ||||      ||| |   |      ||    |  |   ||
Db        895 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
                |||     |||  |    |  |||||||||| ||||| | | ||||||||| |
Db        955 CGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              | ||||| ||  ||||  || |          |||||||| |||||||||  | | |||
Db       1015 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              | ||  ||||  | |      | |   ||||||| || ||  | ||  | |||   |

Db       1075 CGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCG 1134

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGC 931
                ||  | | | |      |||| ||||| ||
Db       1135 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1165


RESULT 13
US-10-363-345A-18588/c
; Sequence 18588, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18588
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18588
US-10-363-345A-18588

  Query Match            12.2%;  Score 122.2;  DB 8;  Length 1172;
  Best Local Similarity   49.2%;  Pred. No. 1.2e-20;
  Matches  458;  Conservative    0;  Mismatches  468;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
              | || |||||   |||    || |||| |  |||   | || |  |     |  |||| |
Db        936 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
              |  |   | | ||| || | |     || |||||| |        || |||||||
Db        876 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 818

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              | ||| ||| |  |  || |||| |||   |  || |       ||| |   |      |
Db        817 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 758

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
                ||||        ||  |||  ||         |      || |     |    ||
Db        757 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 698

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
              ||| | | ||| || || ||| || | |||||| |||||||| |||| || ||||||
Db        697 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 638

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
                ||| |||||  ||||          || ||     | |    |||  |  | |  |
Db        637 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 578

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              || | ||          |||   || | ||| |||     ||  |||    | ||  ||
Db        577 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 518

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
                |||| |||||                |||| ||     ||||    ||  |||   |
Db        517 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 458

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              ||| ||| ||    |  |  | ||  ||||||  ||| | || |   |  |||||
Db        457 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCG 399

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
                || || ||  |||||||  |   |||||  ||    |||       |   ||| |  |
Db        398 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 339
```

```
Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
Db        338 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
Db        278 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 219

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
Db        218 CGCGCGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 159

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
Db        158 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 99

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
Db         98 CGCGCGCGCGCGCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGCGCGCG 39

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGC 931
Db         38 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 8


RESULT 14
US-10-363-483A-18585/c
; Sequence 18585, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18585
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18585
US-10-363-483A-18585

  Query Match             12.2%;  Score 122.2;  DB 9;  Length 1172;
  Best Local Similarity   49.2%;  Pred. No. 1.2e-20;
  Matches  458;  Conservative    0;  Mismatches  468;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
Db        936 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877

Qy         64 CAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAG 123
Db        876 CGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 818

Qy        124 GAGCG-GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
Db        817 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 758

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
Db        757 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 698

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTC-CGCGCGAGCGCGCGCGGGCGC-CGGGCGCGCTG 300
Db        697 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 638

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
Db        637 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 578

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
Db        577 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 518

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
Db        517 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 458

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
Db        457 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-CGCGCGCGCGCGCGCGCGC 399

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
Db        398 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 339

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
Db        338 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
Db        278 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 219

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
Db        218 CGCGCGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 159

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
Db        158 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 99

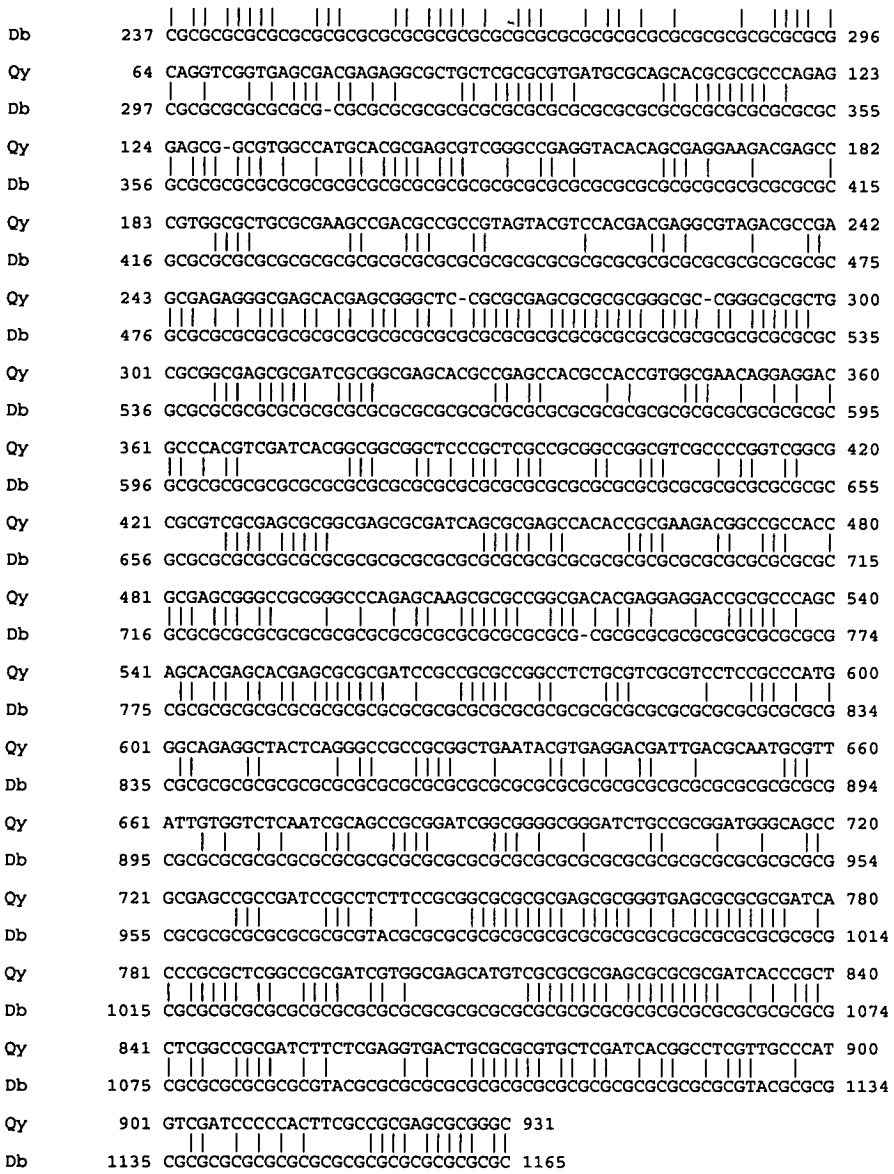
Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
Db         98 CGCGCGCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACGCGCG 39

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGC 931
Db         38 CGCGCGCGCGCGCGCGCGCGCGCGCGCGC 8


RESULT 15
US-10-363-483A-18586
; Sequence 18586, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18586
;   LENGTH: 1172
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 18586
US-10-363-483A-18586

  Query Match             12.2%;  Score 122.2;  DB 9;  Length 1172;
  Best Local Similarity   49.2%;  Pred. No. 1.2e-20;
  Matches  458;  Conservative    0;  Mismatches  468;  Indels    5;  Gaps    5;

Qy          4 CACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAG 63
```



Search completed: April 27, 2006, 14:39:38
Job time : 1004.57 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 628.575 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 cgccacgagcgcgaaggcgg.....accgccggccgcctcgaagc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	93	9.3	9975	7	US-10-496-351-4	Sequence 4, Appli
2	93	9.3	82746	7	US-10-496-351-56	Sequence 56, Appl
c 3	87	8.7	88421	14	US-11-205-109-1	Sequence 1, Appli
4	84.6	8.5	14055	7	US-10-496-351-1	Sequence 1, Appli
c 5	84.6	8.5	82746	7	US-10-496-351-56	Sequence 56, Appl
c 6	84	8.4	14055	7	US-10-496-351-1	Sequence 1, Appli
c 7	79.2	7.9	2247	7	US-10-496-351-8	Sequence 8, Appli
c 8	78.8	7.9	9222	7	US-10-496-351-2	Sequence 2, Appli
c 9	74.6	7.5	3354	7	US-10-496-351-12	Sequence 12, Appl
c 10	74.4	7.4	10968	14	US-11-075-185-35	Sequence 35, Appl
c 11	74.4	7.4	78869	14	US-11-075-185-1	Sequence 1, Appli
c 12	72.6	7.3	4770	7	US-10-496-351-3	Sequence 3, Appli
c 13	72.2	7.2	3305	14	US-11-017-550-68	Sequence 68, Appl
c 14	72.2	7.2	113193	7	US-10-478-943D-1	Sequence 1, Appli
15	69.8	7.0	3354	7	US-10-496-351-12	Sequence 12, Appl
16	69.8	7.0	9222	7	US-10-496-351-2	Sequence 2, Appli
c 17	68.2	6.8	1074	7	US-10-496-351-19	Sequence 19, Appl
18	67	6.7	88421	14	US-11-205-109-1	Sequence 1, Appli
c 19	66.4	6.6	4563	7	US-10-194-487-311	Sequence 311, App
c 20	66.4	6.6	4563	7	US-10-195-883-311	Sequence 311, App
c 21	66.4	6.6	4563	7	US-10-195-888-311	Sequence 311, App
c 22	66.4	6.6	4563	7	US-10-195-889-311	Sequence 311, App
c 23	66.4	6.6	5679	14	US-11-075-185-36	Sequence 36, Appl
24	66.2	6.6	10968	14	US-11-075-185-35	Sequence 35, Appl
c 25	66.2	6.6	11070	14	US-11-075-185-34	Sequence 34, Appl
26	66.2	6.6	78869	14	US-11-075-185-1	Sequence 1, Appli
c 27	66	6.6	1026	7	US-10-496-351-17	Sequence 17, Appl
28	65.6	6.6	908	10	US-10-301-480-592317	Sequence 592317,
29	65.6	6.6	908	10	US-10-301-480-1205726	Sequence 1205726,
c 30	65.6	6.6	963	8	US-10-858-730-182	Sequence 182, App
c 31	65.4	6.5	1011	7	US-10-496-351-46	Sequence 46, Appl
32	65.4	6.5	11905	7	US-10-496-351-57	Sequence 57, Appl
33	64.4	6.4	1074	7	US-10-496-351-19	Sequence 19, Appl
34	64.4	6.4	3408	8	US-10-858-730-40	Sequence 40, Appl
35	63.4	6.3	1326	11	US-11-096-568A-17633	Sequence 17633, A
36	63.4	6.3	4770	7	US-10-496-351-3	Sequence 3, Appli
c 37	63	6.3	11905	7	US-10-496-351-57	Sequence 57, Appl
c 38	62.6	6.3	1618	11	US-11-096-568A-12724	Sequence 12724, A
c 39	62.6	6.3	6413	11	US-11-134-587B-11	Sequence 11, Appl
c 40	62.6	6.3	6604	11	US-11-134-587B-1	Sequence 1, Appli
c 41	62.4	6.2	3167	6	US-09-925-065A-21764	Sequence 21764, A
c 42	62.4	6.2	3167	9	US-10-301-480-123001	Sequence 123001,
c 43	62.4	6.2	3167	10	US-10-301-480-736410	Sequence 736410,
44	62.4	6.2	3886	14	US-11-087-100-37	Sequence 37, Appl
45	62.4	6.2	3886	14	US-11-087-084-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-10-496-351-4/c
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 9975
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-4

Query Match 9.3%; Score 93; DB 7; Length 9975;
Best Local Similarity 46.7%; Pred. No. 1.4e-10;
Matches 466; Conservative 0; Mismatches 520; Indels 11; Gaps 5;

```
Qy          5 ACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGC 64
              ||||||||||     || || ||     ||     | |||  | |  | | | ||  ||||
Db       2504 ACGAGCGCGCGCCCGTCGGCGCTGTCCTGCACCGCGCCGGTGAGCACGGGGTGCGCGCTG 2445

Qy         65 AGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
              |  |||  || || |  | ||| |    ||  ||  | |         ||        ||  |
Db       2444 ACCTCGACGAACGCCCGGTGGCCCGCGTCCAGCAGGGTCGTGGTGGCGGCGTGGAAGCCG 2385

Qy        125 AGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCG 184
              | ||     |  || |  ||||| |       |||  || ||   || ||  | | |||| |
Db       2384 ACCGCGCCGCGCAGGTTCGCGAACCAGTACCCGGCGTCCAGCTCGGGGCCGTCCAGCCAG 2325

Qy        185 TGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGC 244
                ||     | |  |   | |          || ||  | |    |     | |  | ||||||
Db       2324 GCGCCGGTGGTGCTGGACAGGAACGGCACGTGCGCGCGCCGGGGCCGGACGGGGGCGAGC 2265
```

```
Qy        245 GAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCG 304
              | |  ||| || ||     |   |    ||| | |  ||| | ||  ||||   |   ||
Db       2264 GCGGTGGCCAGGACCTCGCGCAGCACGTCGACCCCCGGCGAGTGCGAGGCGTAGT---CG 2208

Qy        305 GCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCC 364
              |  |  ||  |  |||  ||||||||  |   | || || ||||| || |     |  |
Db       2207 ACCGGGACGCGCCGGGCCCGCACGCCGGTCGCGGTCAGCGCGGCGAGCACGGCCTCCACG 2148

Qy        365 ACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCG 424
              |  |||    | |||       |  |   |  |||||  ||| | ||     || |
Db       2147 TCCGCGACGGGGCCGGACAGCACGACGGACTCCGGCC--CGTTGACCGCCGCGACCGCGA 2090

Qy        425 TCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGA 484
              | ||| ||   ||| | ||  | |||  ||    || |  |  |  ||  ||  | ||
Db       2089 GCCCGAACGACTCGACCTCGTGCGGCGGCAGGGCGACGGACAGCATCGCGCCCCGCCCGG 2030

Qy        485 GCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCA 544
              | ||     ||||          ||     | ||| |||||  ||  | ||| ||  |||
Db       2029 ACAGGTGCTCGGCCACCGCGCGGCTGCGCAGCGCCACGACCAGCGCGGCGTCCTCCAGGG 1970

Qy        545 CGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCA 604
              ||||  ||  ||||       |||||| ||  |||  | | |       |||   | ||
Db       1969 TGAGCGCGCCCGCGACGCAGGCCGCGGCGATCTCGCCCTGGGAGTGCCCGAGGACGGCGT 1910

Qy        605 GAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTG 664
              || |   |   |   |||| |  |   |  |  || |      || |||    |
Db       1909 CCGGTTCGACTCCGTGCGCGCGCC--ACAGCCCGGCGAGGGACACGGCGATGGCGAACGC 1852

Qy        665 TGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGA 724
              |    |         |         | || || ||   | |  | ||  ||   || |
Db       1851 GGCGACCTGCACGGTCTCGACCCGGTCGAGGGGGCCGCCGGTGAGGGCGGCGCGCAGGTC 1792

Qy        725 GCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACC-C 783
              | || || |  ||        |||  |||     ||||   |   | |   |||| |||| |
Db       1791 GACGTCGGTGTGCGGGGCGAGCGCGGCGGCGCAGCGGTCGACCGAGTCGGCGAACACCGC 1732

Qy        784 GCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTC 843
              |  ||||| |   | |   |||  |||| ||  |  |   | || |   |     |
Db       1731 GGACTCGGTCAGCAGCCGCGCGCCCATGCCGACCCAGGCGGTGCCCTGGCCGGGGAAGAC 1672

Qy        844 GGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTC 903
              |  |||| || |  ||         | | |      || || | ||  || |  |   ||
Db       1671 GAACGCGGTCCTGCCGTCCACGTCGGCCACCCCGGTGACCAGGTCCGGGTGGGGC---TC 1615

Qy        904 GATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGC 963
              |  | ||     |||  |||||||||| || |||| |   ||| || ||||||  | || |
Db       1614 GCCCGCCGCGAGCGCGTCGAGCGCGGTCCGCGCCCCGGCGCGGTCCGCGGCGACGACCAC 1555

Qy        964 GAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAG 1000
              |  || | |      | |  ||| | |  || ||| |
Db       1554 GGCGCGGTGGGTGAGCCGGGCGCGGCCGTCCGCGAGG 1518
```

RESULT 2
US-10-496-351-56
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 82746
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

Query Match 9.3%; Score 93; DB 7; Length 82746;
Best Local Similarity 46.7%; Pred. No. 8.6e-11;
Matches 466; Conservative 0; Mismatches 520; Indels 11; Gaps 5;

```
Qy          5 ACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGC 64
              |||||||||     || || ||     ||     | |||  | |  | | | | ||  ||||
Db      64697 ACGAGCGCGCGCCCGTCGGCGCTGTCCTGCACCGCGCCGGTGAGCACGGGGTGCGCGCTG 64756

Qy         65 AGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGG 124
              |  |||  || || |  | |||  |   ||  ||  | |  |           ||     || |
Db      64757 ACCTCGACGAACGCCCGGTGGCCCGCGTCCAGCAGGGTCGTGGTGGCGGCGTGGAAGCCG 64816

Qy        125 AGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCG 184
              | ||     | ||  | |||||  |      |||  || ||  || ||  | | |||| |
Db      64817 ACCGCGCCGCGCAGGTTCGCGAACCAGTACCCGGCGTCCAGCTCGGGGCCGTCCAGCCAG 64876

Qy        185 TGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGC 244
              ||     | |  |  | |          || ||  | |   |   |  | | |||||
Db      64877 GCGCCGGTGGTGCTGGACAGGAACGGCACGTGCGCGCGCCGGGGCCGGACGGGGGCGAGC 64936

Qy        245 GAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCG 304
              | |  ||| || ||     |   |    ||| | |  ||| | ||  ||||   |   ||
Db      64937 GCGGTGGCCAGGACCTCGCGCAGCACGTCGACCCCCGGCGAGTGCGAGGCGTAGT---CG 64993

Qy        305 GCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCC 364
              |  |  ||  |  |||  ||||||||  |   | || || ||||| || |     |  |
Db      64994 ACCGGGACGCGCCGGGCCCGCACGCCGGTCGCGGTCAGCGCGGCGAGCACGGCCTCCACG 65053

Qy        365 ACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCG 424
              |  |||    | |||       |  |   |  |||||  ||| | ||     || |
Db      65054 TCCGCGACGGGGCCGGACAGCACGACGGACTCCGGCC--CGTTGACCGCCGCGACCGCGA 65111

Qy        425 TCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGA 484
              | ||| ||   ||| | ||  | |||  ||    || |  |  |  ||  ||  | ||
Db      65112 GCCCGAACGACTCGACCTCGTGCGGCGGCAGGGCGACGGACAGCATCGCGCCCCGCCCGG 65171

Qy        485 GCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCA 544
              | ||     ||||          ||     | ||| |||||  ||  | ||| ||  |||
Db      65172 ACAGGTGCTCGGCCACCGCGCGGCTGCGCAGCGCCACGACCAGCGCGGCGTCCTCCAGGG 65231

Qy        545 CGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCA 604
              ||||  ||  ||||       |||||| ||  |||  | | |       |||   | ||
Db      65232 TGAGCGCGCCCGCGACGCAGGCCGCGGCGATCTCGCCCTGGGAGTGCCCGAGGACGGCGT 65291

Qy        605 GAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTG 664
              || |   |   |   |||| |  |   |  |  || |      || |||    |
Db      65292 CCGGTTCGACTCCGTGCGCGCGCC--ACAGCCCGGCGAGGGACACGGCGATGGCGAACGC 65349

Qy        665 TGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGA 724
              |    |         |         | || || ||   | |  | ||  ||   || |
Db      65350 GGCGACCTGCACGGTCTCGACCCGGTCGAGGGGGCCGCCGGTGAGGGCGGCGCGCAGGTC 65409

Qy        725 GCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACC-C 783
              | || || |  ||        |||  |||     ||||   |   | |   |||| |||| |
Db      65410 GACGTCGGTGTGCGGGGCGAGCGCGGCGGCGCAGCGGTCGACCGAGTCGGCGAACACCGC 65469

Qy        784 GCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTC 843
              |  ||||| |   | |   |||  |||| ||  |  |   | || |   |     |
Db      65470 GGACTCGGTCAGCAGCCGCGCGCCCATGCCGACCCAGGCGGTGCCCTGGCCGGGGAAGAC 65529

Qy        844 GGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTC 903
              |  |||| || |  ||         | | |      || || | ||  || |  |   ||
Db      65530 GAACGCGGTCCTGCCGTCCACGTCGGCCACCCCGGTGACCAGGTCCGGGTGGGGC---TC 65586

Qy        904 GATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGC 963
              |  | ||     |||  |||||||||| || |||| |   ||| || ||||||  | || |
Db      65587 GCCCGCCGCGAGCGCGTCGAGCGCGGTCCGCGCCCCGGCGCGGTCCGCGGCGACGACCAC 65646

Qy        964 GAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCGAAG 1000
              |  || | |      | |  ||| | |  || ||| |
Db      65647 GGCGCGGTGGGTGAGCCGGGCGCGGCCGTCCGCGAGG 65683
```

```
RESULT 3
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Staffa, Alfredo
;  TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
;  FILE REFERENCE: 3002-2US
;  CURRENT APPLICATION NUMBER: US/11/205,109
;  CURRENT FILING DATE:  2005-08-17
;  PRIOR APPLICATION NUMBER: US/09/976,059
;  PRIOR FILING DATE: 2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 88421
;   TYPE: DNA
;   ORGANISM: Actinoplanes sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2077)..(3078)
;   OTHER INFORMATION: ORF 1; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3118)..(4032)
;   OTHER INFORMATION: ORF 2; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4038)..(5048)
;   OTHER INFORMATION: ORF 3; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6665)..(5814)
;   OTHER INFORMATION: ORF 4; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7703)..(6693)
;   OTHER INFORMATION: ORF 5; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9464)..(8130)
;   OTHER INFORMATION: ORF 6; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9691)..(10761)
;   OTHER INFORMATION: ORF 7; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (12751)..(10829)
;   OTHER INFORMATION: ORF 8; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13617)..(12802)
;   OTHER INFORMATION: ORF 9; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15203)..(13614)
;   OTHER INFORMATION: ORF 10; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15591)..(15863)
;   OTHER INFORMATION: ORF 11; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15880)..(19035)
;   OTHER INFORMATION: ORF 12; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19032)..(39713)
;   OTHER INFORMATION: ORF 13; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (39713)..(65800)
;   OTHER INFORMATION: ORF 14; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65826)..(66530)
;   OTHER INFORMATION: ORF 15; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (66546)..(67370)
;   OTHER INFORMATION: ORF 16; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (67384)..(70059)
;   OTHER INFORMATION: ORF 17; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70099)..(70662)
;   OTHER INFORMATION: ORF 18; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70659)..(71906)
;   OTHER INFORMATION: ORF 19; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (73439)..(71964)
;   OTHER INFORMATION: ORF 20; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (74216)..(73563)
;   OTHER INFORMATION: ORF 21; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75424)..(74213)
;   OTHER INFORMATION: ORF 22; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75535)..(76464)
;   OTHER INFORMATION: ORF 23; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78110)..(76449)
;   OTHER INFORMATION: ORF 24; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79864)..(78107)
;   OTHER INFORMATION: ORF 25; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81624)..(79861)
;   OTHER INFORMATION: ORF 26; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81909)..(81682)
;   OTHER INFORMATION: ORF 27; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82346)..(82062)
;   OTHER INFORMATION: ORF 28; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82587)..(84446)
;   OTHER INFORMATION: ORF 29; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84481)..(85548)
;   OTHER INFORMATION: ORF 30; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85556)..(86845)
```

```
;   OTHER INFORMATION: ORF 31; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87372)..(86803)
;   OTHER INFORMATION: ORF 32; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87494)..(88420)
;   OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

  Query Match             8.7%;  Score 87;  DB 14;  Length 88421;
  Best Local Similarity   46.2%;  Pred. No. 1.4e-09;
  Matches  442;  Conservative    0;  Mismatches  500;  Indels   15;  Gaps    4;

Qy         12 CGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGG 71
Db      66801 CGGCCGCCGCGGTGAGCGCCGCGAGGTCGCGCACGTCCGCCCGGACGGTGAGCACCGCGG 66742

Qy         72 TGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCG 131
Db      66741 CGCCCTGCTTGCGGCAGAGGTCGGCGGTGTGCTCCAGCTGGCCCACGGTGCCCAGCGGAT 66682

Qy        132 TGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCT 191
Db      66681 AGGGCACCCCCGGCAGGTCGGCCGCGCGGTCC------AGGAGCACCAGGTCGGCTCCCT 66628

Qy        192 GCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGG 251
Db      66627 CGGCGGCGAAGGCCACGGCGCAGGCGCGGCCGAGGCCGCGGGCCGCGCCGGTGACGATGG 66568

Qy        252 CGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCG 311
Db      66567 CGGTCTTGCCGGTCAAGCGCACGGGTTCTCCTCGGGTTCAGTCACGGTCCAGGTCGGCGG 66508

Qy        312 CGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGA 371
Db      66507 CGATCAGGCGCAGCAGGCCGGCCGGGTCCTCGGTGAGGTACATGTGGCCGCCGTCCACCT 66448

Qy        372 TCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAG 431
Db      66447 CGAC-----CAGCTGGAACGGGCCGCTGGTGGCCTTGCTCCACCCGGCCGTCTGCTCGGC 66393

Qy        432 CGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCC 491
Db      66392 GGAGACCAGCGCGTCGTCGCGGGCGTGCAGCGCGGTGACGGGTGCGGGCAGCGGATCGGC 66333

Qy        492 GCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGAC---CGCGCCCAGCAGCACGAG 548
Db      66332 GCTGCCCGCCACGTAGCTCTCGTGCATCTCGACGTCGGCGCGCAGCGCGGGCAGGATCAG 66273

Qy        549 CACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGAGG 608
Db      66272 CTCGCGCATGTCCGGATCGGCGAGCGCCTCGTGGTCGTAACCGGCGAACTCGCGCACCCG 66213

Qy        609 CTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGT 668
Db      66212 CAGCAGGAACTCCTCGTCGTCCAGGCCGCTCGCCCGGTTGGTGCGA-GGGGTCCAGGGAT 66154

Qy        669 CTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCG 728
Db      66153 CGGGCGAGCCGCTGACGAAGAGCCGCTCCACGCCCACCTCGCCGTCGCGCACCAGCGCGT 66094

Qy        729 CCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCT 788
Db      66093 GCGCTATCTCGTACGCGAGGACGGCGCCGAGGCTGTGCCCGAAGAGCGCCACCCGGCCCC 66034

Qy        789 CGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCG 848
Db      66033 CCGCGCCCACCTCGTCGGCGACGGTGAGCGCCAGGTCCGCGGCCGCCGCGACGACGTCCG 65974

Qy        849 CGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGATCC 908
Db      65973 CGTACGGTTCGTCGAGCAGCCGCCACTCCCGGCCCGGAAGCTGGAGCGCGACGATCGGCC 65914

Qy        909 CCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGA 965
Db      65913 GGTCCGGCCCGGCGAGCTCGGCCCACGGGTGGAAGAACGAGGCGCCGGCACCGGCGA 65857


RESULT 4
US-10-496-351-1
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             8.5%;  Score 84.6;  DB 7;  Length 14055;
  Best Local Similarity   45.6%;  Pred. No. 6.6e-09;
  Matches  453;  Conservative    0;  Mismatches  529;  Indels   11;  Gaps    4;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
Db       7614 CGACCTGCGCGACGCGGGCCCGGACCGCACCGCCGAACTGCTCGCCGAGGTGCTGGGGCT 7673

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
Db       7674 GTTCGCGGCGGGCCGCCTTGAGCCGCTGCCGCTGACCGTCCGCGACGTCCGGCACGCCCC 7733

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
Db       7734 CGAGGCGCTGCGGCACGTCGCGTCCGGCGGGCACGTCGGCAAGGTCGTGCTCACCGTCCC 7793

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
Db       7794 CCGCCCGCTCGACCCCGAGGGCACCGTCCTGGTCCACGGCGGCGGCGCGCTCGGCGCGCT 7853

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
Db       7854 CACCGCCCGCCACCTGGTGTCCGCGCACGGCGCCCGCCACCTGCTGATCACCTCCCGGCG 7913

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
Db       7914 CGGCCCCGACGCGCCCGGCGCCGCCGAGCTGGCCGCCGACCTCGCCGAACTCGGCGCCAC 7973

Qy        363 C---CACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGC 419
Db       7974 CACGCGGTTCGAGCGCTGCGACGCCGCCGACCGCGACCGGGTCGCCGCCCTGCTGGCCGA 8033

Qy        420 GCGCGTCGCGAGCGCGGCGAGCGCGAT----CAGCGCGAGCCACACCGCGAAGACGGCCG 475
Db       8034 CCTGGGCGGGCGCCCGCTGACCGCCGTCGTGCACACGGCGGGCGCGCTCGACGACGCCCT 8093

Qy        476 CCACCGCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGC 535
Db       8094 GCTCGCCGACCAGGACGCGGCGCGCGTCGACGCGGTCCTGCGCCCCAAGCTCGACGGGGC 8153

Qy        536 CCAGCAGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGC 595
Db       8154 GCTGCTGCTGGACGAGCTGACCCGCGACGCCGACCTGGCCGCGTTCGTGCTGTTCTCCTC 8213

Qy        596 CCATGGGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAAT 655
```

```
Db       8214 CGCCGCCGGGACGCTCGGCAACGCGGGCCAGGCCAACTACGCCGCCGCGAACGCCGCGCT 8273

Qy        656 GCGTTATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGG 715
              ||      ||| | |    |     | | |   ||| |  ||  | || |  ||  ||||
Db       8274 -CGACGCGCTGGCCGCCCGCCGCCGCTCCGCCGGGCTGCCCGCCACCTCCATCGCCTGGG 8332

Qy        716 CAGCCGCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGC 775
                | |   | |      ||||    |||| | || |   | ||| |  || | |||||
Db       8333 GCCACTGGGAGCACACCAGCGCCCTCACCGCCGGGCTC---ACCGCCGCCGACCTCGCGC 8389

Qy        776 GATCACCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCAC 835
              |      ||  || ||||  |          ||  |     |  ||
Db       8390 GCGGGGCCAGGCACGGCGCGGGCTCGCTCACCACCACGAGGCGCTGGCCCTGCTCGACG 8449

Qy        836 CCGCTCTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTG 895
              |||| |||| |        | ||  ||    ||  || |||||| | |  ||| ||  |
Db       8450 CCGCGCTCGCCTCGCCGGAACCCGCCGTCCTCGCCGCCGCGCTCGACCCCCGCCGCGCCG 8509

Qy        896 CCCATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCG 955
              || | |||   || |  || |||  |     ||| || |||||  | |     | |   |
Db       8510 CCGAGGTCCCGCCGCTGCTGCGCGACCTCGCCGGACCGCGCCGGGCGGTCGCGGCCCAGG 8569

Qy        956 TGGGCCGCGAGGCCGGGCCATGCCGGACCGCCG 988
                | |  ||  || |  |     | || || ||
Db       8570 CCGTCGCCGGCGCGGACCGGCCGCTGATCGACG 8602


RESULT 5
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
;   LENGTH: 82746
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

  Query Match             8.5%;  Score 84.6;  DB 7;  Length 82746;
  Best Local Similarity   45.6%;  Pred. No. 4.4e-09;
  Matches  453;  Conservative    0;  Mismatches  529;  Indels   11;  Gaps    4;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              | ||  |||||| |  |||   || ||| ||| ||     |      || | |    |
Db      29428 CGACCTGCGCGACGCGGGCCCGGACCGCACCGCCGAACTGCTCGCCGAGGTGCTGGGGCT 29369

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              |   |  | |  || | ||| ||||||     |   |      |   || | |
Db      29368 GTTCGCGGCGGGCCGCCTTGAGCCGCTGCCGCTGACCGTCCGCGACGTCCGGCACGCCCC 29309

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              |||  ||   | || |      |  ||  ||  |   |  || |    ||  ||
Db      29308 CGAGGCGCTGCGGCACGTCGCGTCCGGCGGGCACGTCGGCAAGGTCGTGCTCACCGTCCC 29249

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              |    ||||   |   | | ||| | | ||  |      | | |   | | |||
Db      29248 CCGCCCGCTCGACCCCGAGGGCACCGTCCTGGTCCACGGCGGCGGCGCGCTCGGCGCGCT 29189

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |   || | |  | |   || |   |||  |||   | | |     | | | |||
Db      29188 CACCGCCCGCCACCTGGTGTCCGCGCACGGCGCCCGCCACCTGCTGATCACCTCCCGGCG 29129

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              ||||      ||  | |||||     ||     | |   | |||  | |||||  |  | | |
Db      29128 CGGCCCCGACGCGCCCGGCGCCGCCGAGCTGGCCGCCGACCTCGCCGAACTCGGCGCCAC 29069

Qy        363 C---CACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGC 419
              |   |  |||| | | ||| || | ||  |       |||   ||  | |  | | | |
Db      29068 CACGCGGTTCGAGCGCTGCGACGCCGCCGACCGCGACCGGGTCGCCGCCCTGCTGGCCGA 29009

Qy        420 GCGCGTCGCGAGCGCGGCGAGCGCGAT----CAGCGCGAGCCACACCGCGAAGACGGCCG 475
              | | || | || ||  || ||| |    ||   | |       | ||| |||| ||
Db      29008 CCTGGGCGGGCGCCCGCTGACCGCCGTCGTGCACACGGCGGGCGCGCTCGACGACGCCCT 28949

Qy        476 CCACCGCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGC 535
              | |  ||| | || ||||  |  |   | |||  || ||| | |  ||   ||| | ||
Db      28948 GCTCGCCGACCAGGACGCGGCGCGCGTCGACGCGGTCCTGCGCCCCAAGCTCGACGGGGC 28889

Qy        536 CCAGCAGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGC 595
              | || ||  ||  |  |  |      | |  |  | |  | |||   || |||| |
Db      28888 GCTGCTGCTGGACGAGCTGACCCGCGACGCCGACCTGGCCGCGTTCGTGCTGTTCTCCTC 28829

Qy        596 CCATGGGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAAT 655
              |    |  || |||  ||  || | |  |||  | ||||      |||  | |||  |
Db      28828 CGCCGCCGGGACGCTCGGCAACGCGGGCCAGGCCAACTACGCCGCCGCGAACGCCGCGCT 28769

Qy        656 GCGTTATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGG 715
              ||      ||| | |    |     | | |   ||| |  ||  | || |  ||  ||||
Db      28768 -CGACGCGCTGGCCGCCCGCCGCCGCTCCGCCGGGCTGCCCGCCACCTCCATCGCCTGGG 28710

Qy        716 CAGCCGCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGC 775
                | |   | |      ||||    |||| | || |   | ||| |  || | |||||
Db      28709 GCCACTGGGAGCACACCAGCGCCCTCACCGCCGGGCTC---ACCGCCGCCGACCTCGCGC 28653

Qy        776 GATCACCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCAC 835
              |      ||  || ||||  |          ||  |     |  ||
Db      28652 GCGGGGCCAGGCACGGCGCGGGCTCGCTCACCACCACCGAGGCGCTGGCCCTGCTCGACG 28593

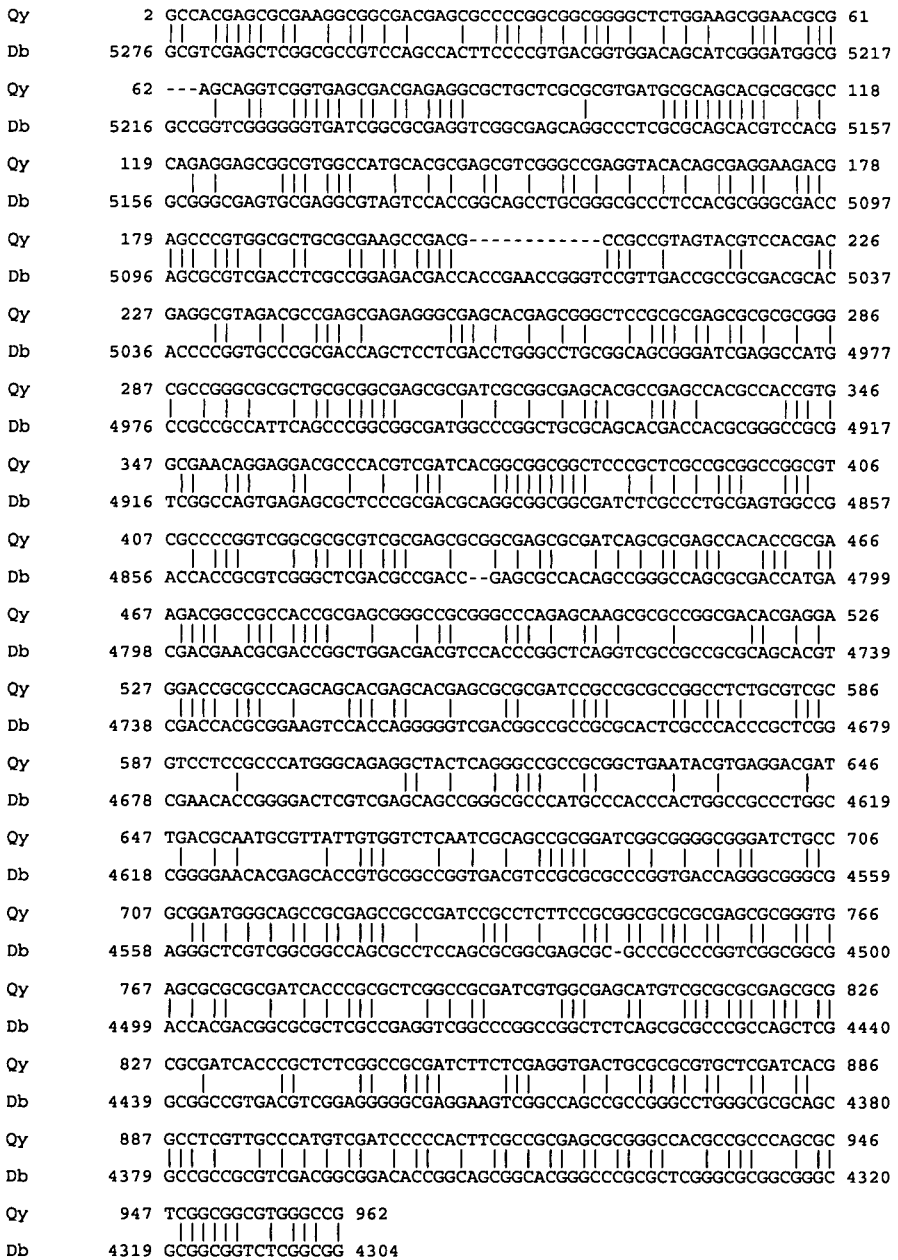
Qy        836 CCGCTCTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTG 895
              |||| |||| |        | ||  ||    ||  || |||||| | |  ||| ||  |
Db      28592 CCGCGCTCGCCTCGCCGGAACCCGCCGTCCTCGCCGCCGCGCTCGACCCCCGCCGCGCCG 28533

Qy        896 CCCATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCG 955
              || | |||   || |  || |||  |     ||| || |||||  | |     | |   |
Db      28532 CCGAGGTCCCGCCGCTGCTGCGCGACCTCGCCGGACCGCGCCGGGCGGTCGCGGCCCAGG 28473

Qy        956 TGGGCCGCGAGGCCGGGCCATGCCGGACCGCCG 988
                | |  ||  || |  |     | || || ||
Db      28472 CCGTCGCCGGCGCGGACCGGCCGCTGATCGACG 28440


RESULT 6
US-10-496-351-1/c
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             8.4%;  Score 84;  DB 7;  Length 14055;
  Best Local Similarity   45.9%;  Pred. No. 8.8e-09;
  Matches  448;  Conservative    0;  Mismatches  510;  Indels   18;  Gaps    4;
```



RESULT 7
US-10-496-351-8/c
; Sequence 8, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2247
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-8

Query Match 7.9%; Score 79.2; DB 7; Length 2247;
Best Local Similarity 47.9%; Pred. No. 1.3e-07;
Matches 298; Conservative 0; Mismatches 313; Indels 11; Gaps 2;

RESULT 8
US-10-496-351-2/c
; Sequence 2, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz

```
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 9222
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-2

  Query Match             7.9%;  Score 78.8;  DB 7;  Length 9222;
  Best Local Similarity   49.7%;  Pred. No. 1.1e-07;
  Matches  289;  Conservative    0;  Mismatches  282;  Indels   11;  Gaps    3;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGGCTCTGGAAGCGGAACGCGA 62
              | |  || |||  | ||||| ||      || | |||    || | ||  |||| ||  |
Db       2704 CGAGCAGGGCGGTGACGGCGTCGGCGAACCGGACGGGCTCGCGCAGGTTGCGGTACCAGT 2645

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
              ||  | ||  ||||   | |  | || |  ||     | ||  |||||   |   |  |
Db       2644 GCTCGGCGTCGAGCCGGGTGGCGGGCCGACCGGTGACGGTGGACAGCAGCGGGATCTCGG 2585

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              |      | | |||  | |     ||| |  |     || | | |    |||  | ||| |
Db       2584 GTTCGCGGGCGGCGACGTCGCCGAGTGCGGCCAGCAGCTCCGCGCGCAGGGCGTCGACCT 2525

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              ||   ||| ||  |  | || |||||      |  ||||  |      |      | ||
Db       2524 GCGGGGAGTGCCCGGCGAAGTCGACGCCGGGCAGCCGCCACCGCAGCACCCGCCGGGCG- 2466

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              || ||| |  || ||| |  || |  |||   ||||  |  ||    | ||     ||
Db       2465 --GACAGGACCCGCTCGAACTCGCGCACCGC---CGCGTCCTCGCCGGAGACGGTGACCG 2411

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              ||| || || |  ||||   |||||  | || | ||  |      ||   || ||  ||
Db       2410 ACGCGGGCCCGTTGACGGCCGCCACGCACACCCGCCCCTCGAACCCGGCGAGCAGCGCGG 2351

Qy        363 CCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCG 422
              ||       ||| |||||  |  | |   ||  ||  | ||| |    | |   ||| |
Db       2350 TGACCCGCTCGACGTCGGCGACGACGGTGAGCATCGCGCCGCGGCCGGACAGCTCGCGGG 2291

Qy        423 CGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGC 482
              ||    | |||  |||     || ||   ||| ||   | ||   | ||  ||  | |||
Db       2290 CGATCAGCGCGGAGCG-----GAGCACGACGATCCGGGCGGCCTCGGCGAGCGTGAGCGC 2236

Qy        483 GAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAG 542
              |  || | || ||||    |  ||       || ||||      || |||||||  |  || |
Db       2235 GCCCGCGACGTGGGCGGCGGCCACCTCGCCCTGCGAGTGCCCGACGACCGCGTCGGGCCG 2176

Qy        543 CACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTC 584
              ||||   ||   ||| |    |||||||   | | |  || |
Db       2175 CACGCCGACCGACGCCCAGAGCGCCGCGAGCGACACGACGAC 2134


RESULT 9
US-10-496-351-12/c
; Sequence 12, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 3354
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-12

  Query Match             7.5%;  Score 74.6;  DB 7;  Length 3354;
  Best Local Similarity   48.3%;  Pred. No. 1e-06;
  Matches  245;  Conservative    0;  Mismatches  254;  Indels    8;  Gaps    1;

Qy         19 GGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTGAGCGA 78
              |||| ||| |   |   ||| |  | |  ||  || |    || ||  | ||| |   |
Db       3192 GGCGGCGACCTTGCGCTCGGAGCCGGTGAGGCCGCTGGTGTCGTGCTCGACGGGGTCGGC 3133

Qy         79 CGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTGGCCAT 138
              || |    | | |  ||||| |    || |          | ||        | |  |  |
Db       3132 CGCGTCCGGTTCCAGGCGCGCGCACAGCGGTCGCGCGCCACACGCCCTGGCCAGGTGCCA 3073

Qy        139 GCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGCGCGAA 198
              |  |  | ||  |  | || ||          | ||  | ||||  ||||| ||   |||
Db       3072 GGCCCTGCGCAGCACGGCGCGGGCGCGGCGGTGGTCGCCGAGGGCGTGGTGCGCGGCGCC 3013

Qy        199 GCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCAC 258
              | |  | ||  |  |    | |  |     ||  ||   ||| | ||  ||||||    |
Db       3012 GAGGCAGGCGAGGGTGGCGGCCAGCTCGTAGCGGTCACCGCAGCGCTCCAGGGCGTCGGC 2953

Qy        259 GAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGATCGC 318
              | |  |    ||  | ||| |||  ||  | | | ||   ||| ||||||||||  |
Db       2952 GGCCTCGGCGAGCAGGTGCGGGCGGCGGGTCGGCGGGCTGAGCGGGGCGAGCGCGCGCAG 2893

Qy        319 GGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCACGGC 378
              || ||| | | | |   || |  ||||      | ||  |  ||||| |||
Db       2892 CGCCAGCGCCCTGGGGCGGCCCGCGTCGGCGCCGGTCGTGGCGAGCTGGTCGAACAC--- 2836

Qy        379 GGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGCGCGGCG 438
              |  ||||  | ||||| ||  || |||| ||  ||  | || | ||  ||  ||  ||
Db       2835 -----CAGCCGCCGGGCGCGGTCGTGGTTGCCCAGGCGGGTCCACGCCTCGGCGGCCCCG 2781

Qy        439 AGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGCGGGCC 498
              |  |||||| |    |  | || | ||  |        ||  || ||| | |
Db       2780 GTGCGCCACGGCGCGAGGCCGGGCAGGTCGAGGCCCCACGAGCGCAGGAGGTCGCCGCAC 2721

Qy        499 CAGAGCAAGCGCGCCGGCGACACGAGG 525
              || |||  |||  ||| | || ||
Db       2720 GCCAGGAAGTCCGCGAGCGCCGCGTGG 2694


RESULT 10
US-11-075-185-35/c
; Sequence 35, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
;   LENGTH: 10968
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-35
```

```
  Query Match             7.4%;  Score 74.4;  DB 14;  Length 10968;
  Best Local Similarity   47.3%;  Pred. No. 8.5e-07;
  Matches  329;  Conservative    0;  Mismatches  356;  Indels   11;  Gaps    3;

Qy        134 GCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGC 193
              || ||  |||||    | | |    |      || |||  | ||   ||| || ||  |
Db       4898 GCGATCGGCGCGATGCCCTGCCGTCTGAGGCGTGCCAGGTCGGCGGCGCCGAGGTGCGCC 4839

Qy        194 GCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCG 253
              |  | |||||  ||| | |  | |    |  | |||||  |         ||  | | ||
Db       4838 GTCATGCCGAGCCCGGCCTGCTCCCAGAAGCCCCAGGCGAGGCTCATCGCCGGCACGCCG 4779

Qy        254 AGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCG 313
              || || |  |   | |||||||||| | ||  |  | |  |   |||  |  |||
Db       4778 CGCGCGCGGAGCTGCGCCGCGAGCGCGTCCAGGAACGCGTTGGCCGCGGCGTAGTTCGCC 4719

Qy        314 ATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGA-- 371
              | | |||   |       |  ||| |||  | |||||| | || | || ||
Db       4718 TGCCCCGCGCTGCCCAGCGTGCCGGCCGCCGACGAGAACAGCACGAAGGCCGCGAGCCGA 4659

Qy        372 -----TCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTC 426
                  || || ||  ||   || | |  | || |||| ||      | |             |
Db       4658 TGGCCTCGCGTCGCCGCGGCCAGGTGCCAGCCGCCGTCGATCTTCGGCCGCAGCACGCGC 4599

Qy        427 GCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGC 486
              ||||||    ||  || || | ||  |     ||  ||    |  | || ||  | ||
Db       4598 GCGAGCTGCTCGGCCGAGAGCTTCGTCACGATGACGTCGTCCAGCACGCCCGCCATGTGC 4539

Qy        487 GGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACG 546
              | |    |   | |      |||||   | |   || |||||| ||||||  | ||
Db       4538 ACGACCGCCGTGAGGGGGTGTGCAGCCGGGATCGCCGCGATGACCGCCCCCAGCGCCGCG 4479

Qy        547 AGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGA 606
              ||    ||     |||||| || |  |  |||||| |  ||||     |
Db       4478 CCGTCGGCGACGTCGCACGCCGCGACGTCGAC--CGTCGCGGCGCCCGCGGCGCGCAGCT 4421

Qy        607 GGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTG 666
              |    ||| ||||| ||| | |        |  ||    ||  || |     |
Db       4420 CGTCCACCAGCGCCGCGGCGTCCGGCGCATCCATCCCGCGCCGCGACGTGAGCACGAGGT 4361

Qy        667 GTCTCAATC--GCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGA 724
              | | ||  |     ||||||    || |  ||| || |||||||    |  |  | | |
Db       4360 GCCGCACGCCGTGCGCCGCGACCAGGTGCCGCGCGACCTGCCGCCCGAGCTCCCCGGTGC 4301

Qy        725 GCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCG 784
              | |      | | |   | || || |  | |||   |  |  |||||    ||  |||
Db       4300 CGCCCGTGATCAGCACCGTCCCGCCGGGATCCAGCCGCGTCGGTCGCGCCGCGTCGGCCG 4241

Qy        785 CGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCG 820
              |||  ||  |||       ||  || || ||||||||
Db       4240 CGCCCGCGCGCACCTCGCGCAGGCGTGGCGCGCGCG 4205


RESULT 11
US-11-075-185-1/c
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 78869
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-1

  Query Match             7.4%;  Score 74.4;  DB 14;  Length 78869;
  Best Local Similarity   47.3%;  Pred. No. 5.4e-07;
  Matches  329;  Conservative    0;  Mismatches  356;  Indels   11;  Gaps    3;

Qy        134 GCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGC 193
              || ||  |||||    | | |    |      || |||  | ||   ||| || ||  |
Db      20165 GCGATCGGCGCGATGCCCTGCCGTCTGAGGCGTGCCAGGTCGGCGGCGCCGAGGTGCGCC 20106

Qy        194 GCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCG 253
              |  | |||||  ||| | |  | |    |  | |||||  |         ||  | | ||
Db      20105 GTCATGCCGAGCCCGGCCTGCTCCCAGAAGCCCCAGGCGAGGCTCATCGCCGGCACGCCG 20046

Qy        254 AGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCG 313
              || || |  |   | |||||||||| | ||  |  | |  |   |||  |  |||
Db      20045 CGCGCGCGGAGCTGCGCCGCGAGCGCGTCCAGGAACGCGTTGGCCGCGGCGTAGTTCGCC 19986

Qy        314 ATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGA-- 371
              | | |||   |       |  ||| |||  | |||||| | || | || ||
Db      19985 TGCCCCGCGCTGCCCAGCGTGCCGGCCGCCGACGAGAACAGCACGAAGGCCGCGAGCCGA 19926

Qy        372 -----TCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTC 426
                  || || ||  ||   || | |  | || |||| ||      | |             |
Db      19925 TGGCCTCGCGTCGCCGCGGCCAGGTGCCAGCCGCCGTCGATCTTCGGCCGCAGCACGCGC 19866

Qy        427 GCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGC 486
              ||||||    ||  || || | ||  |     ||  ||    |  | || ||  | ||
Db      19865 GCGAGCTGCTCGGCCGAGAGCTTCGTCACGATGACGTCGTCCAGCACGCCCGCCATGTGC 19806

Qy        487 GGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACG 546
              | |    |   | |      |||||   | |   || |||||| ||||||  | ||
Db      19805 ACGACCGCCGTGAGGGGGTGTGCAGCCGGGATCGCCGCGATGACCGCCCCCAGCGCCGCG 19746

Qy        547 AGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGA 606
              ||    ||     |||||| || |  |  |||||| |  ||||     |
Db      19745 CCGTCGGCGACGTCGCACGCCGCGACGTCGAC--CGTCGCGGCGCCCGCGGCGCGCAGCT 19688

Qy        607 GGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTG 666
              |    ||| ||||| ||| | |        |  ||    ||  || |     |
Db      19687 CGTCCACCAGCGCCGCGGCGTCCGGCGCATCCATCCCGCGCCGCGACGTGAGCACGAGGT 19628

Qy        667 GTCTCAATC--GCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGA 724
              | | ||  |     ||||||    || |  ||| || |||||||    |  |  | | |
Db      19627 GCCGCACGCCGTGCGCCGCGACCAGGTGCCGCGCGACCTGCCGCCCGAGCTCCCCGGTGC 19568

Qy        725 GCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCG 784
              | |      | | |   | || || |  | |||   |  |  |||||    ||  |||
Db      19567 CGCCCGTGATCAGCACCGTCCCGCCGGGATCCAGCCGCGTCGGTCGCGCCGCGTCGGCCG 19508

Qy        785 CGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCG 820
              |||  ||  |||       ||  || || ||||||||
Db      19507 CGCCCGCGCGCACCTCGCGCAGGCGTGGCGCGCGCG 19472


RESULT 12
US-10-496-351-3/c
; Sequence 3, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
```

```
;   LENGTH: 4770
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-3

  Query Match             7.3%;  Score 72.6;  DB 7;  Length 4770;
  Best Local Similarity   47.9%;  Pred. No. 2.4e-06;
  Matches  240;  Conservative    0;  Mismatches  259;  Indels    2;  Gaps    1;

Qy          1 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 60
              |||| || || ||| ||||| ||| || | || |||           |   | |  ||
Db       3693 CGCCGCGTGCACGACGGCGGTGACCTGCTCGCCCGCGCGCTCCAGGCCGGTCAGCAGGGC 3634

Qy         61 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 120
              |   ||  ||| | |    | || |        ||   ||| ||     |  ||| |  |
Db       3633 GGCGAGCGCGGCGCGGTCGGCGACGTCGCAGGCGGCGGTGGTGACGCGGGCGCCCAGCGC 3574

Qy        121 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 180
                ||   | ||| |      |          ||| |  |   |   | ||  |  |||
Db       3573 CTCCAGGTCGGCGGCGAGCGCGGGCGCGCCGGGGGCCGAGGGGCCGCGGCGGCTGGTGAG 3514

Qy        181 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 240
              || || | |  ||| |  | || ||  |  |  | |     || ||||  | ||||
Db       3513 GACGAGGTGGTCGGCGCCGAGGGCGGCGAGGCGGCGGGCCACGTGCGCGGCGACGCCGCC 3454

Qy        241 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 300
              |     |  || || || || || ||  |||| |  |    | | |  | ||||  || |
Db       3453 GGTTCCGCCGGTGATCAGGACCGTGCCGCGCGGGGTCCACGGGGCGGCCGGGGCCTGCGG 3394

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
              |||||  || ||| |  |      |   ||  |  |||     ||||| | ||     |
Db       3393 TTCGGCGCGCACGAGCCTGCGCGCGAACACGCCCTGCGCGCGGATGGCGACCTGGTCCTC 3334

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              ||  |||||   ||| ||  ||  |       | ||  | ||    ||| || | |
Db       3333 ACCTCCGTCGTCGACGCCG--GGTCCGGCGAGGACGGCGGCGAGCCCGCGGCGCTGGCGC 3276

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              |    ||| |  |||||||  ||| |||| ||  |||      ||  |    ||  | |
Db       3275 GGGTCGGCGGGGTCGGCGAGGTCGACCAGCCCGCCCCAGCGGCCGGGGTGCTCCAGCGCG 3216

Qy        481 GCGAGCGGGCCGCGGGCCCAG 501
              |||   ||||  |  |||||
Db       3215 ACGACGCGGCCCAGTCCCCAG 3195


RESULT 13
US-11-017-550-68/c
; Sequence 68, Application US/11017550
; Publication No. US20050250183A1
; GENERAL INFORMATION:
;  APPLICANT: The Scripps Research Institute
;  APPLICANT:  Schultz, Peter G
;  APPLICANT:  Wang, Lei
;  APPLICANT:  Anderson, John C
;  APPLICANT:  Chin, Jason
;  APPLICANT:  Liu, David R
;  APPLICANT:  Magliery, Thomas
;  APPLICANT:  Meggers, Eric L
;  APPLICANT:  Mehl, Ryan A
;  APPLICANT:  Pastrnak, Miro
;  APPLICANT:  Santoro, Stephen  W
;  APPLICANT:  Zhang, Zhiwen
;  TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
;  FILE REFERENCE: 54-000120US
;  CURRENT APPLICATION NUMBER: US/11/017,550
;  CURRENT FILING DATE:  2004-12-17
;  PRIOR APPLICATION NUMBER: US/10/126,927
;  PRIOR FILING DATE: 2002-04-19
;  PRIOR APPLICATION NUMBER: US 60/285,030
;  PRIOR FILING DATE: 2001-04-19
;  PRIOR APPLICATION NUMBER: US 60/355,514
;  PRIOR FILING DATE: 2002-02-06
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68
;   LENGTH: 3305
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-11-017-550-68

  Query Match             7.2%;  Score 72.2;  DB 14;  Length 3305;
  Best Local Similarity   45.0%;  Pred. No. 3.1e-06;
  Matches  438;  Conservative    0;  Mismatches  523;  Indels   12;  Gaps    4;

Qy          2 GCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCG 61
              ||  || | |||||   | |  |||  || ||     |  ||   ||  | ||  ||||
Db       2389 GCGGCGGGGGCGAACATCGGGTTGAGGCCCACGTGCTGGACGCCGGGGGCGTGGGCCGCG 2330

Qy         62 AGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAG 121
              |||    |||  |    ||   ||      | |       | |||||  ||| ||  |
Db       2329 AGCTCCGCGGCCATGCCCGTCCGGACGGACAGGGTGTCGGCGAGCAGCGCGCCCGGTCGC 2270

Qy        122 AGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGC 181
              | |||| | ||  |        | ||  | |  || | | | || | |    |  | ||  ||
Db       2269 ATGAGCCGGGTCACGGGCGCCACGGCCTTGAGGGCCACGTCCTCGTGTACGGCGAGCAGG 2210

Qy        182 CCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCG 241
              || ||  |    |  | || |  ||  | | || || || |  | |  | ||||
Db       2209 ACGAGGTCCGCGTCCCGGAGGGCGGCCGCGAGCTCGGGCCCCGGCGCGGTGACGTCGCCC 2150

Qy        242 AGCGAGAGGGCGAGCACGAG---CGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGC 298
              | |  | ||||  |    |  |||   | |  ||||  ||     | |||  |  ||
Db       2149 ACCAGGCAGGCGTCCGGCCGTCCCGGCGGCGGTACGAGGTCGACGACGAGCGTGCGGCTG 2090

Qy        299 TGCGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGG 358
              |||  |  ||   | | |||||| || ||||  |  |||  |  || ||   | || |
Db       2089 CCCGCCTCCCGCAGCAGCCCGGCGAACATGCCGCCCACCGCTCCGCTGCCGCCGACGACG 2030

Qy        359 ACGCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGG 418
              |||| |  |  ||  ||     | | | |  |   || | ||||     || |||
Db       2029 ACGCTCCGGGGGAAGCCGCTCATCGGGCGCCCGCCACTGCGCTGCCGTCG---AGGGCGG 1973

Qy        419 CGCGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCA 478
              |  | | ||| | ||   || | || || ||  || |  | || |   | |||
Db       1972 TGACCATGGCGCGGGCCTTGACCACGGTCTGCCTGAACTCCTCCTCCTGGTCGGAGAGGG 1913

Qy        479 CCGCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCA 538
              | ||| ||  |||| | | | || |  |  |          | || |  ||  | ||  |
Db       1912 ACACGATCGCCCCGCCGACGCCGAACTCGGCCCGGCCGTCGGCCAGCACGATGGTGCGGA 1853

Qy        539 GCAGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCA 598
              |  | |  | |  |||          |  || |  | ||  |  ||          |
Db       1852 TGACGATGCTGAGGTCGGCGGCGCCGCTGAGGGCGAACCATCCGAGCGCCCCGGGTAAGA 1793

Qy        599 TGGGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCG 658
              |  | | ||    ||      || | | |||    |   || || ||    |
Db       1792 CGCCCCGGGGGCCTTCCTCCAGGCGGTCGATGATCTCCATGGGTCGCTTCTTGGGCGCGC 1733

Qy        659 TTATTGTGGTCTCAATCGCAGCCGCGGATCG-GCGGGGCGGGATCTGCCGCGGATGGGCA 717
              |  |||  |   |   ||||  ||  ||  |  ||   ||| |  |  ||||
Db       1732 CGGTCATGGAGCCGCCGGGGAAGGCGGCGCGTACGCAGGCGGCGGTGCTGGTGCCGGGCC 1673

Qy        718 GCCGCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGA 777
              || ||   || | |||  |  | ||  |  |  | | | ||  | | | | | | |
Db       1672 GCAGCCGTCCCCGGATGGTCGACACCAGCTG-----GTGCACGGGCGCGAGGTCTCCCAC 1618

Qy        778 TCACCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCC 837
              |  ||  || | ||      || |   ||  |||   ||      | |       |
Db       1617 CTCGAAGAGCCGGGGCACGTGGACGGAGCCGATCGCGCAGACGCTGTTGAGGTCGTTGCG 1558

Qy        838 GCTCTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCC 897
              |  | | |  ||||| || |   |    ||  | | |||     |||||  || | |
Db       1557 GACCAGGTCGACGATCATCAGGTTCTCGGCCCGGTCCTTCTCCCGGCCGGCCAGGTCGGC 1498

Qy        898 CATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTG 957
```

```
Db       1497 GCGGAGCCGCTCGTCCTCCTCCGCCGGTGCGCCCCGGGGGCGGGTCCCCTTGATGGGCTT 1438

Qy        958 GGCCGCGAGGCCG 970
              || | ||| ||||
Db       1437 GGACTCGACGCCG 1425


RESULT 14
US-10-478-943D-1/c
; Sequence 1, Application US/10478943D
; Publication No. US20060073574A1
; GENERAL INFORMATION:
;  APPLICANT: University College Dublin, National University of Ireland
;  TITLE OF INVENTION: Engineered biosynthesis of novel polyenes
;  FILE REFERENCE: PC-1636US
;  CURRENT APPLICATION NUMBER: US/10/478,943D
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: S2001/0527
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 113193
;   TYPE: DNA
;   ORGANISM: Streptomyces nodosus
US-10-478-943D-1

  Query Match             7.2%;  Score 72.2;  DB 7;  Length 113193;
  Best Local Similarity   46.0%;  Pred. No. 1.4e-06;
  Matches  460;  Conservative    0;  Mismatches  503;  Indels   36;  Gaps    5;

Qy          3 CCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGA 62
              ||  || |||   ||||| |         |||  ||||      | ||||   |  |   |
Db      87477 CCGGGATCGCCGGGGCGGAGTACGAGAGCCCCTCGGCCACCGACCGGAACTCGTCCAGCA 87418

Qy         63 GCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGA 122
               | ||||  | |||| |||| ||      | ||    |    || |   |     ||
Db      87417 TCGGGTCCATCAGCGGCGAGTGGAAGGCGTGGCTGACGCGCAGCCGGCTGGTGCGCCGTC 87358

Qy        123 GGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCC 182
              ||   |||       |  |||| ||   || ||  |  | | |||  | |||||| |
Db      87357 CCTGCTCGGTGAAGTGCCCGGCGACCGCGAGGGCGGCGGCCTCGGCGCCGGAGACGACCA 87298

Qy        183 CGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGA 242
              ||      | |  ||  | ||| || ||  |                ||| ||| | ||  ||
Db      87297 CGGACGACGGTCCGTTGACGGCGGCGATG----------------GAGACGTCGCCGGTGA 87253

Qy        243 GCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCG 302
              |     ||  || | ||  |  | |   | |||  |||  |  |   |  | |     ||
Db      87252 GGAGCGGGAGGACCTCGTCCTCGGTGGCCTCGACGGCGATCATGGCGCCGCCCTCGGGCA 87193

Qy        303 CGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGC 362
              ||  ||    |  || ||| |  || |||              | | |   ||||  ||
Db      87192 GCGCCTGCATCAGGGCCGCGCGGGCGGCGACGAGACGGCAGGCGTCCGCCAGGGTGAGCA 87133

Qy        363 CCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCG 422
              || || ||||    |||||||||      ||| |||  |  |  | || ||  |  | |  |
Db      87132 CCCCGGCGATGTGCGCGGCGGCGAT--CTCACCGATGGAGTGGCCGGCCACGAAGTCCGG 87075

Qy        423 CGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGC 482
              | || |   |  ||    | | |   |  ||||  |||| ||| | ||  |    | |
Db      87074 CTTCACACCCCAGGACTCCACCAGTCGGTAGAGCGCCACCTCGACGGCGAACAGTGCGGG 87015

Qy        483 GAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAG 542
              | |  |  | || |  |  ||             || |  |  | ||| | |||  ||
Db      87014 CTGGGTCCATCCGGTCTCGTTCA-----------GCAGGTCGGCGTCCTCACCCCACAT 86967

Qy        543 CACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGG 602
              |||   |  ||||  |  |||  | ||||| ||    || || || |  || ||| ||
Db      86966 CACCTCCCGCAGCGAACCCTCCAACTCGCCGTCCAGGACGGCGAGTACGGCGTCCAGGGC 86907

Qy        603 CAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTAT 662
```

```
Db      86906 CTCGGTGAACACGGGGACCCTGCCGTACAGCTCACGGCCCATGCCGAGGCGCTGGGAGCC 86847

Qy        663 TGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGC 722
              |    | |    |||||  |||| |    ||      |  |   |    ||
Db      86846 CTGGCCGGAGAACAGGAACGCGGTGCGGCCGCCGCGGGCGGTGCCGGTCAGCAGCGCGCT 86787

Qy        723 GAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACC 782
              |       | | ||  |    || ||| | ||| | | | |     | |||       |
Db      86786 GTCGTCGGTACCTGCGGCAAGGGCCGCGAGGGCGCG-GAGCGCCTCGGTGCGGTCGGCGC 86728

Qy        783 CGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCT 842
              || |  |||| |||      | |||  | |  |||   |       ||||  || || | |
Db      86727 CGAGGACGGCGGCGCGGTGGTCGAAGACGGAGCGGCTGGTGAGCAGCGAGTACGCGATGT 86668

Qy        843 CGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCA------CGGCCTCGTTGC 896
              |||| | |  | ||||| |||        | ||  ||| ||      |||||| || ||
Db      86667 CGGCGGGGCGCGTCTCGGGGTGGGCCTCGACATGGGCGAGCAGTCGGGCGGCCTGGTCGC 86608

Qy        897 CCATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCCAGCGCTCGGCGGCGT 956
              || | |          ||||||  |  ||||        || |||    | ||  | |||
Db      86607 GCAGGGCCTCGGGGGTCTTCGCGGACAGCGGCCAGGGGGCGGCCGGAAGTTCCTCCGCGG 86548

Qy        957 GGGCCGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCT 995
              || |      || ||| |    ||||| ||| |||||| ||
Db      86547 GGCCGCTCGGGGCGGCCGGGGCCGGGCCGGCGGCCGGCT 86509


RESULT 15
US-10-496-351-12
; Sequence 12, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 3354
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-12

  Query Match             7.0%;  Score 69.8;  DB 7;  Length 3354;
  Best Local Similarity   45.8%;  Pred. No. 9.6e-06;
  Matches  348;  Conservative    0;  Mismatches  407;  Indels    4;  Gaps    3;

Qy        188 CGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAG 247
              ||| | | |  || ||  |   |||      | ||        || ||    || || | |
Db       1166 CGCCGGGGGGCGCTGAGCCGCTCGTGCCCGGCCCGGTGTCTGGGGTGAGGGCTGATCCCG 1225

Qy        248 AGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCG 307
              || | | |||  |    ||  ||  || | |   |||     ||   | | ||  |
Db       1226 CCGGGGGGGACGGGGCTTCCGCGTTCGGTCGGGTGGCTGCGGATCACGTCGTTCCCGCTG 1285

Qy        308 AGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACG 367
              |  |    | | ||| |  | | || ||    |    | ||| | ||  | || ||
Db       1286 GGGTCCGCCCCCGCGCGTCCTCGGACCCCGCCGCGCCCGTCGACCTGGTCGCCGACCTGC 1345

Qy        368 TCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCG 427
              ||| ||  | |  || ||||   ||| ||  ||| ||| |   ||  |  || ||  | |
Db       1346 TCGGTCGGGCCCCCGCCTCC--GTCGTCGACGCCCGCGCCCGCCTCGCGGGAGCCGGGCT 1403

Qy        428 CGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCG 487
              ||| || | ||| | ||          |      | || ||| | |  |||||
Db       1404 CGACCGGGCCGATCCCGCCGCTGTGCTGGACGTGCTGCCCGCCGCCCGGCGCGCCGAGCT 1463
```

```
Qy        488 GGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGA 547
              |  |  | || | | ||   ||   |     | ||  |||   | ||||       |||
Db       1464 GCACCTGCGCGCCGCGCGGACGTTGCACGAGCGCGGTGAGTCGCCCGCCGTCGTCGCCGA 1523

Qy        548 GCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGAG 607
               |||     | |   | | |   |||  | |  |   ||   |  ||  |  |
Db       1524 CCACCTCGTCACCGCCGGGCTGCCCGCGCCCGCCTGGGCCGTCGACGTGCTCGTCGCCGC 1583

Qy        608 GCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGG 667
                 |   ||   ||| || ||  ||||   ||    | ||||  ||         |
Db       1584 CGCCGACCACGCGCTCGCCGCCGACGACGTGCCCACCGCCGTCGCACGGCTCGCCGCCGC 1643

Qy        668 TCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCC 727
               | |   ||   |  ||          |||||  | || | |   |    ||  || ||
Db       1644 CCACCGCCGGGACGACGCGGACCAACCGCGGGCCGTCCCCCCGCGCGCAGTCCCGGACCC 1703

Qy        728 GCCGATCCGCCTCTTCCGCGGCGCGCGCG-AGCGCGGGTGAGCGCGCGCGATCACCCGCG 786
              |||  ||  |  |  |||  ||| | |||  |  |||  |  ||||| ||   | |  ||
Db       1704 GCCCGTCGCCGCCCACCGGTGCGGGTGCGCGGACCGGACGGCCGCGCCCGCCGAGCGCCG 1763

Qy        787 CTCGGC-CGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGG 845
              | || | | |||    | ||      | || || || || | |       | ||| | ||
Db       1764 CACGTCGCCCGAGGCCGCCGCCGTCGCCGGGCCCGCGCTCACCGCCCGGCTCGCCCGCGC 1823

Qy        846 CCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGA 905
               |||   |  ||| |   | ||||     ||  | | | || ||   | ||| | |
Db       1824 GCGCTGGCAGCTCAACCCGGCTGCCGCGCTGTGCCACCTCGACCCGCTCGCCGACGCGCT 1883

Qy        906 TCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGC 944
                || |    | |  || || ||  || ||   | | ||
Db       1884 CGCCGCCGACCCCGACGCGCTCGCCCCGCGGGACGCCGC 1922
```

Search completed: April 27, 2006, 15:52:58
Job time : 629.575 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 4961.09 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 cgccacgagcgcgaaggcgg.....accgccggccgcctcgaagc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	137.6	13.7	2243	10	AG381986	AG381986 Mus muscu
2	130.6	13.0	1798	10	AG171124	AG171124 Pan trogl
3	127.2	12.7	1347	10	CL496012	CL496012 SAIL_619_
4	126	12.6	1516	2	BG809984	BG809984 mgct002xd
5	125.6	12.5	1878	10	AG393083	AG393083 Mus muscu
6	125.6	12.5	1981	10	CL515231	CL515231 SAIL_898_
c 7	120.2	12.0	1798	10	AG171124	AG171124 Pan trogl
8	119.6	11.9	1409	2	BG850730	BG850730 1024029B1
c 9	119.6	11.9	1542	10	AG032943	AG032943 Pan trogl
c 10	118.8	11.9	1516	2	BG809984	BG809984 mgct002xd
11	118.6	11.8	1452	10	AG032979	AG032979 Pan trogl
12	118.2	11.8	1280	10	CL470317	CL470317 SAIL_140_
13	117.6	11.7	1126	10	AG064051	AG064051 Pan trogl
c 14	117.6	11.7	1299	10	AG039481	AG039481 Pan trogl
c 15	117.2	11.7	1452	10	AG032979	AG032979 Pan trogl
c 16	116.6	11.6	1278	10	AG060116	AG060116 Pan trogl
17	116.2	11.6	1542	10	AG032943	AG032943 Pan trogl
18	116	11.6	1309	10	AG077201	AG077201 Pan trogl
c 19	116	11.6	1462	10	AG441877	AG441877 Mus muscu
c 20	115.8	11.6	1123	10	AG080476	AG080476 Pan trogl
21	115.2	11.5	1341	10	AG030611	AG030611 Pan trogl
22	115.2	11.5	1625	10	AG043477	AG043477 Pan trogl
23	115	11.5	1462	10	AG441877	AG441877 Mus muscu
c 24	115	11.5	1772	10	CL508812	CL508812 SAIL_802_
25	115	11.5	1780	10	AG448243	AG448243 Mus muscu
26	114	11.4	1299	10	AG039481	AG039481 Pan trogl
27	113.8	11.4	937	10	AG154114	AG154114 Pan trogl
28	113.8	11.4	1660	10	AG058676	AG058676 Pan trogl
c 29	113.6	11.3	1409	2	BG850730	BG850730 1024029B1
c 30	113.2	11.3	1165	10	AG030649	AG030649 Pan trogl
31	113.2	11.3	1598	10	AG030579	AG030579 Pan trogl
c 32	113	11.3	1538	10	AG030607	AG030607 Pan trogl
33	113	11.3	1853	10	AG382854	AG382854 Mus muscu
c 34	112.6	11.2	1280	10	CL470317	CL470317 SAIL_140_
c 35	112	11.2	1569	10	AG341503	AG341503 Mus muscu
c 36	112	11.2	1878	10	AG393083	AG393083 Mus muscu
37	111.8	11.2	1538	10	AG030607	AG030607 Pan trogl
38	111.8	11.2	1569	10	AG341503	AG341503 Mus muscu
39	111.8	11.2	1772	10	CL508812	CL508812 SAIL_802_
c 40	111.6	11.1	1780	10	AG448243	AG448243 Mus muscu
41	111.4	11.1	1131	10	AG042920	AG042920 Pan trogl
c 42	111.2	11.1	1281	2	BG852363	BG852363 1024034A0
43	111.2	11.1	1552	10	AG430101	AG430101 Mus muscu
44	111	11.1	1189	10	AG030608	AG030608 Pan trogl
45	109.8	11.0	1821	10	CL090560	CL090560 ISB1-17N1

ALIGNMENTS

RESULT 1
AG381986
LOCUS AG381986 2243 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-192D21.TJ, genomic survey sequence.
ACCESSION AG381986
VERSION AG381986.1 GI:47993191
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 2243)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .2243

```
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-192D21.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             13.7%;  Score 137.6;  DB 10;  Length 2243;
  Best Local Similarity   47.9%;  Pred. No. 2.3e-15;
  Matches  478;  Conservative    0;  Mismatches  501;  Indels   18;  Gaps    5;

Qy          1 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 60
              |||   | ||  |  |||||   || | |  |  || | || ||    |  ||    |
Db       1169 CGCGGGGGGCCGGCGGGCGGGCGCGCGGGNGCGCGCCGGGGCGCGGGCGGCGCCCGCCCG 1228

Qy         61 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 120
              |  |  |  || | ||| ||   |   || ||||   |  ||||  |     |||||
Db       1229 GGCCGCGGGGGNGCGCGCCGGNGCGGCGCGCGCGCGGNGGGNGCGCGCCCGCGCCGCCCG 1288

Qy        121 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 180
              |  ||||||  | |  || ||   ||| | ||| | |       | | ||  | || |
Db       1289 CGGCCGCGGCGGCGGCGCGCGCGGCCGCGCCCGGCGGCGCGCGCGCGGGCGGGGGGCGCG 1348

Qy        181 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 240
              |  | |||||   |||    || || ||| |  |   | |      || ||||      ||
Db       1349 CGGGCGGCGCGCGGCGGGCNCGGCGGCGCGGGGGGCGGGCGGGCGCGCGGCGC-----CC 1403

Qy        241 GAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTG 300
              |   |   || || ||      |||  |||| | |||||| |||   | |||    |
Db       1404 GCCGGGCGGGCCGGGCGGCGCGCGGCGNCGCGGGGGCGCGCCCGGCGCCGGGGGCGNCGC 1463

Qy        301 CGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGAC 360
              || ||    |||   |||| | |  ||      |  | |   ||   | ||| |
Db       1464 GGCNGCCGCGGCGGGGGCGGGGCGGGGGCGCGCGCGGGGCGGGGGCGCCGGCGGGGGCGG 1523

Qy        361 GCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCG 420
              | | |     | |||  | ||  |  |   | |||| |||||| |||   |   ||
Db       1524 CCGCGGGGGCGCCGCGGGCGGGGGGCGGGGGGGGCGCGCCCGGCGGCGCGGGGCGGGGGC 1583

Qy        421 CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACC 480
              |  | ||||| || |||||    |||       | |||||  |   ||||||
Db       1584 GCGCGCCGGGGCGCGCCGCGCGCGCCNCGCGGNGCGGNCGCCGCGCGGCGCGCCGCCGCG 1643

Qy        481 GCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGC 540
              |||  |||  ||| |||    |  |  ||||| ||||    ||   |  |  | |||   ||
Db       1644 GCGCCCGGCGCGCCGGC---GGCCGGGCGCGGCGGCCCGGCGGCGGCGCGCNCGCGNGGC 1700

Qy        541 AGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATG 600
              |  ||  | |         |||||||    |||     ||| || | |  | | ||  |
Db       1701 GCGGGGGGCGGGCG-------GCCGCCGCCGGGGCGCGCGCGCCGGGGCGGCGGGCCGGG 1753

Qy        601 GGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTT 660
              ||    | |  |   | | ||||||||  |    ||   |  ||     ||    |||
Db       1754 GGGGCGGNCGGCGGCGCGGCGCCGCGGGCGGGCGCGNCGGCGCGGCGCGCGGCGCGCGGC 1813

Qy        661 ATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCC 720
              | |    |  ||| |  ||       |||||||||    | | ||| |  ||||  |
Db       1814 -GGGNGCCGGCGGCCGCCGGNGCCCGCGCGCGGGGCGGCGGCGGGCGCCGGCGGGCGCCG 1872

Qy        721 GCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCA 780
              ||  || || |  || | |  | |||| | ||||  |||  |    ||    |||  |
Db       1873 GCNCGCGGCGGCCGCGGCNCGGCGGCGGNGNGCGCCGGCGGCGCGCGGCCGCGGCG--CG 1930

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              |||||  ||| ||||  ||  | | ||  | ||||||||  |||||||||     | |
Db       1931 GGCGCGCGCGGGCGCGCGCGGCGGGCGCNCGCCGCGCGCGCGCGCGCGCGGCGCGCNGGN 1990

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              |  || | |        |  | || | || ||  |   | | ||  |  |  |
Db       1991 GNGCGGCGGGCGCGCGCNCGGCGCGGGGNGCNGGCGCGCGCGCGGCGGCGCGGCGGNCNG 2050
```

```
Qy         901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
               |    |  |    |  |||   ||| ||      ||| |  |    |  || | |
Db        2051 GNGNCGCGNCNGNCNCNGCGCCGCCGCCGGGNNGCGCGCGCGCCCGCGGCNGGGGNGNCG 2110

Qy         961 CGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCG 997
               |||| |    || |    ||| |   |||| ||  ||
Db        2111 CGCGCGCNGNGGGCNGCGCGGNCGCGCGGCGGCGCCG 2147


RESULT 2
AG171124
LOCUS       AG171124               1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match             13.0%;  Score 130.6;  DB 10;  Length 1798;
  Best Local Similarity   46.7%;  Pred. No. 4.4e-14;
  Matches  469;  Conservative    0;  Mismatches  519;  Indels   17;  Gaps    7;

Qy          2 GCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTC-TGGAAGCGGAACGC 60
              | | || |||||   |||    || |||| |  |||||| ||| |  |   |||  |||
Db        700 GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGCGCGGCGCGCGCGCGCGCGNCGC 759

Qy         61 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 120
              |  |  |   |   |   |  | |||  || ||| |     |||| ||    ||||
Db        760 GCNCGCGCGCGCCGGNCGCCCGCCGCGGCGCCCGCNCNCCGCGCGCCGCCGCGGCGCNNN 819

Qy        121 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAG-ACGA 179
              | |   ||    ||  || ||   |  || | || |     | |||  |  |  ||
Db        820 GCGCGCCGCGCNCGCGCCGCGCGNCGCCCCCGCGGCGCGCNCGCCCGCGNCGCGCGC 879
```

```
Qy        180 GCCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACG---ACGAGGCGTAGA 236
Db        880 GCNCGCCGCGCCGCGCGCCNCGCGCCCGCGCGNNCCGCGGCGCCGCGCCCGCCNCGCCCG 939

Qy        237 CGCCGAGCGAGAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGC 296
Db        940 CNCGGCGCNCNNCGCCGCCCCGCGCCNCNCGCCGCCGCGGCGCNCGCCGCCCGCGCGCCG 999

Qy        297 GCTGCGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGA 356
Db       1000 CCGCCGCCCNCGGCGCGCCNGCNGCNCGCCCGCGCGCCNNNNNNGCGCCGGCGCGCGCGC 1059

Qy        357 GGACGCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTC 416
Db       1060 GCGNGCGCGCGCGCGCNNNCGCGCCCGCGNCNCNCCGCGNCGCCCGCCCGCGCGCGCGNC 1119

Qy        417 GGCG---CGCGTCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGC 473
Db       1120 CGCGGGCCGCCNCNCGCCCGCCGCGCNNCCGCNCNGCGGCGGCGCGCGCGCGCGCGCGNC 1179

Qy        474 CGCCACCGCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGC 533
Db       1180 GGCGCGCGCCGGCGCGCNCGCGGCNNCGCCCGCCGGCGCCGCCGCGCCGCGCGGCGNCGC 1239

Qy        534 GCCCAGCAGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCC 593
Db       1240 GCGCGGGCGCGCNCGCGCCCGGCNGNGCGCGCNCCCGCNCGCGCCGGCCGCGCCCCCGCG 1299

Qy        594 GCCCATGGGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCA 653
Db       1300 CGCNCGCGGCGCGCGCGCGCGCGGGCCGCGCGCGCGCGGCGCGGGGCGCCGGCGCGCGCG 1359

Qy        654 ATGCGTTATTGTGGTCTCAATCGCAG---CCGCGGATCGGCGGGGCGGGATCTGCCGCGG 710
Db       1360 GCGCGCGGCGCCGCGCGGCGCCGCGGGGNCGCCGCNCCGGCGCCGCGCGCGCCNCCGCGC 1419

Qy        711 ATGGGCAGCCGCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCG 770
Db       1420 NCGGNCCGCGGCGGNCGNCCGGCCGCCNCNCGGCNCGCGCGGCGCCGCGGCGCGCGCGCC 1479

Qy        771 CGCGCGATCACCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCG 830
Db       1480 GGCGCCNCGCCGCNCCNCNCCGCGCCCGCNNCNCNNCCGNNCCGCGCGCCNNCGCGNCCG 1539

Qy        831 ATCACCCGCTCTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCT 890
Db       1540 CGNCNCGCNNCNCNCCGCGCGGCGGCGCGCCGGCGCCGCCGGCGCCCGNCCGCNCGGC-- 1597

Qy        891 CGTTGCCCATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGG 950
Db       1598 ---GCNCCGCCCCGCGCGNCNCGCGCGNCGNNCGCGCGCGCGCCGCCGCCGCGCGCGG 1654

Qy        951 CGGC-GTGGGCCGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCC 994
Db       1655 CGNCGGCCGGCCGCNCGCGCGCGCGCCNCCGCGCCGCCGCCGCCC 1699
```

```
RESULT 3
CL496012
LOCUS       CL496012                1347 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_619_B02.v3 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_619_B02.v3, genomic survey sequence.
ACCESSION   CL496012
VERSION     CL496012.1  GI:45987773
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1347)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
  PUBMED    12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS826533; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1347
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_619_B02.v3"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN
```

Query Match 12.7%; Score 127.2; DB 10; Length 1347;
Best Local Similarity 48.2%; Pred. No. 1.9e-13;
Matches 475; Conservative 0; Mismatches 484; Indels 27; Gaps 4;

```
Qy         20 GCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTGAGCGAC 79
Db        262 GCGGGGCGGGCCGGGCGNCGGGGGCCGCGGGGCCGCCGCCGNNCGCGCCCGCCGCGCCGG 321

Qy         80 GAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTGGCCATG 139
Db        322 GGGCGGCGCGCCGGGGGCGGGGCGGGCCGGGGGGCCGGCCGCCGGCGCCCGGACGCGGCG 381

Qy        140 CACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGCGCGAAG 199
Db        382 GCCGCGCGCGACCCGCCGCGGCGCGCCCCGCGGCGCGCGGGCGCCCGGGCGGCGCGGCCG 441

Qy        200 CCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCACG 259
Db        442 CGGCCGCGCGCGCCCGGCGCGCGCGGGCGCGCGGCGGCGCCGCGCGGCCCCGCGCCCGCG 501

Qy        260 AGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGC--------- 310
Db        502 CGCCGCCGCCGCGGGCGCGCGCGCGCGGCCGCCGGCGGCGGCACGGCGCGCGAGCGGGCG 561

Qy        311 ----GCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCAC 366
Db        562 CGCGGCGCGCGCGCGCGCGGCGCGGAGCCGCGCCCGCCGCGCGCGCGGGCCGACGCGCGG 621

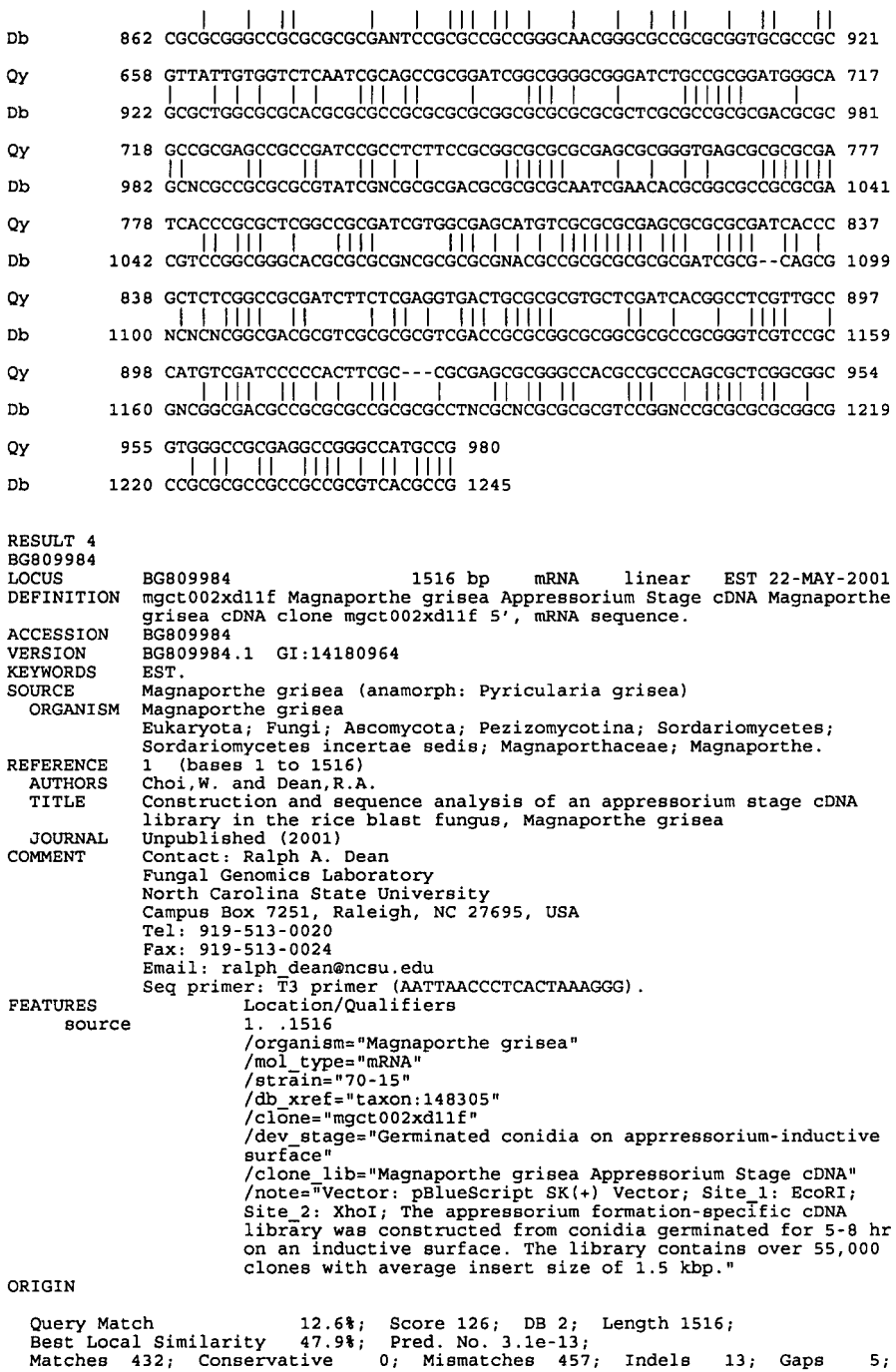
Qy        367 GTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTC 426
Db        622 GCGTCGCGCGTGGNCGTCGCCGCGTCGGGCCGCTCCGGCGCGCGCGGCGCGCGCCGCGGC 681

Qy        427 GCGAGCGCGGCGAGCGCGATCAGCGCGAGC---------CACACCGCGAAGACGGCCGCC 477
Db        682 GCGCTCGGGCGCCCCGCCTCCGCGGCGCGCCGGCGCCGGCGGACGCCGCCGCGGGCCGCG 741

Qy        478 ACCGCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCC 537
Db        742 CGCGCGCGCGCGACCCGCGGCCGCCGCGTGCCGGCCCGGCGCGCCGGCGCGGCGCGGGCG 801

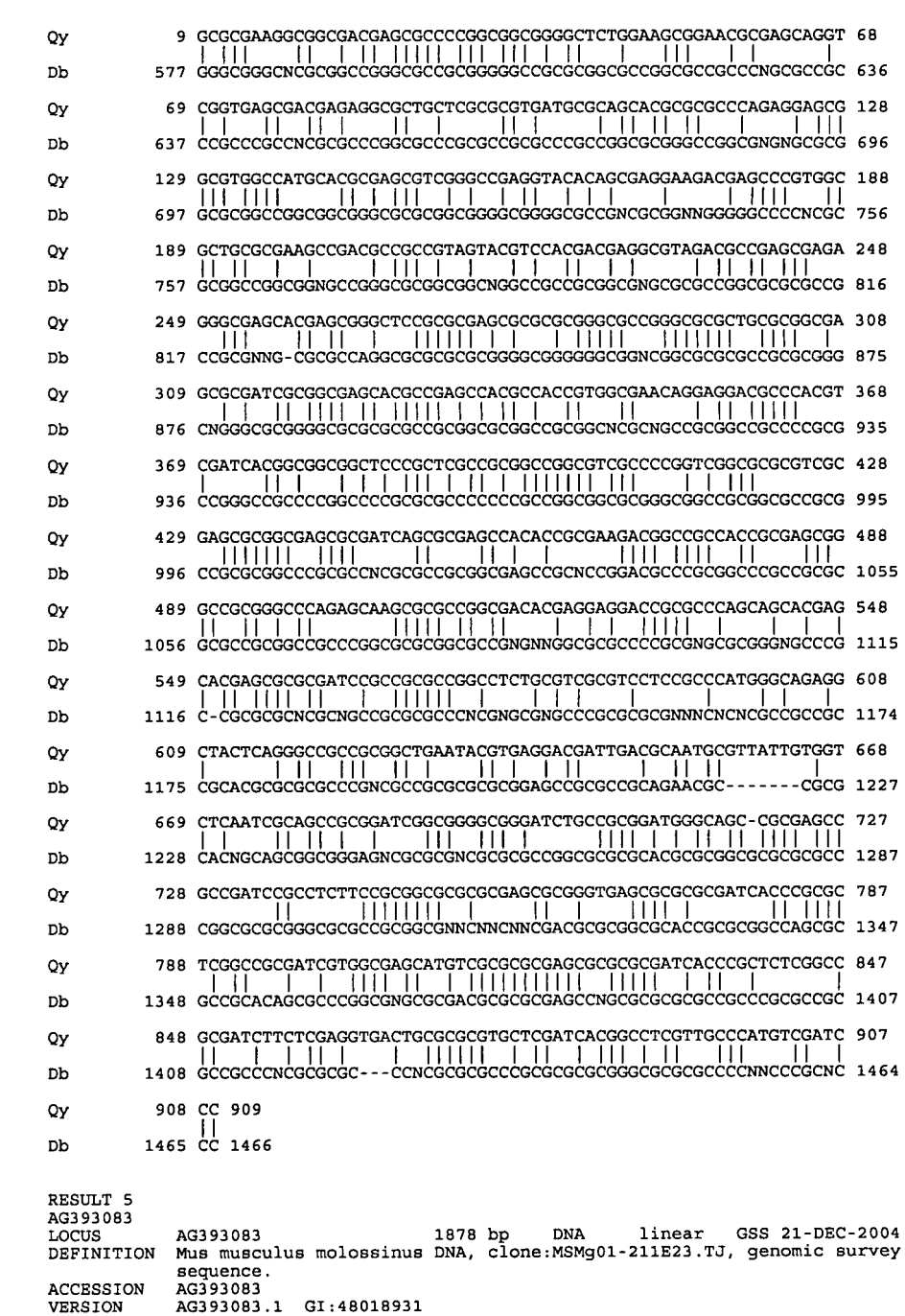
Qy        538 AGCAGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCC 597
Db        802 TCCGCGACGGCCGCGCGCGCCGGCGGCGCGCCGCGCGCGGCGCGCGCGCGCTCGCCGGCG 861

Qy        598 ATGGGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGC 657
```



RESULT 4
BG809984
LOCUS BG809984 1516 bp mRNA linear EST 22-MAY-2001
DEFINITION mgct002xd11f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe grisea cDNA clone mgct002xd11f 5', mRNA sequence.
ACCESSION BG809984
VERSION BG809984.1 GI:14180964
KEYWORDS EST.
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 1516)
AUTHORS Choi,W. and Dean,R.A.
TITLE Construction and sequence analysis of an appressorium stage cDNA library in the rice blast fungus, Magnaporthe grisea
JOURNAL Unpublished (2001)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu
Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES Location/Qualifiers
source 1. .1516
/organism="Magnaporthe grisea"
/mol_type="mRNA"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgct002xd11f"
/dev_stage="Germinated conidia on apprressorium-inductive surface"
/clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
/note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI; Site_2: XhoI; The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library contains over 55,000 clones with average insert size of 1.5 kbp."
ORIGIN

Query Match 12.6%; Score 126; DB 2; Length 1516;
Best Local Similarity 47.9%; Pred. No. 3.1e-13;
Matches 432; Conservative 0; Mismatches 457; Indels 13; Gaps 5;



RESULT 5
AG393083
LOCUS AG393083 1878 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-211E23.TJ, genomic survey sequence.
ACCESSION AG393083
VERSION AG393083.1 GI:48018931

```
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1878)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1878
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-211E23.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match            12.5%;  Score 125.6;  DB 10;  Length 1878;
  Best Local Similarity  48.0%;  Pred. No. 3.5e-13;
  Matches  467;  Conservative    0;  Mismatches  487;  Indels   18;  Gaps    6;

Qy          9 GCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGT 68
              | |||  ||||| |  | |   | |||| ||| |   |  |  |||   ||   |  ||
Db        905 GGGCGGCGGCGGGGCGGGGGCGCGCGGCCGCGCGCGGCGCGGGGCGCGGCGGCGGGCGGC 964

Qy         69 CGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCG 128
               || | | |  | | |   | |          |  |||  |  |  |||| | |  |   |
Db        965 GGGCGCGGGCGGGGGGCGCCGGGGGCGCGCGGCGGCGGCGGGCCNGCGCGCGGGCGGCGG 1024

Qy        129 GCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGC 188
               || ||    ||  ||| |||  | | ||      |    | ||    || || ||  ||
Db       1025 CCGCGGGNCCGCCGGCGGGCGCGGCGGCGGNCGGCCGGCGGGGGCGCGCGGGCGCGGCGC 1084

Qy        189 GCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGA 248
               ||   |||  |  |  || |      |  || |  ||  | ||||  |  | || | |
Db       1085 GCGNGGCGGCGGGGCGGCGGGGCGGGCGCGGCGGCGGGGGGGCGGCGCGGNCGCGGGGCG 1144

Qy        249 GGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGA 308
               |  | ||    |      |||||||  || ||||| ||| |||          ||||||
Db       1145 CGCGGCGCGGCGGGCNCNGCCGCGCGGCCGGGCGCGCGCGGCGGCG-------GCGGCGG 1197

Qy        309 GCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGT 368
```

```
               |||||  || |||| |  ||| | ||  |||   || ||      ||       ||
Db       1198 GCGCGCGCGGGGCGCGGCCGCGGCGCGGCGCGGGCGCGGGCGGGCGGCCCCGCGCCGGCG 1257

Qy        369 CGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGC---GCGT 425
               ||  | |||||| |     ||| || ||  | |||||   | | ||  |||||   |||
Db       1258 CGGNCGCGGCGGGGCNCGGCGCGCGGCGGCGGCGGCGCGCCGCGGGGGGGCGCGGGGCGC 1317

Qy        426 CGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAG 485
              |||| ||||| || |         |||||      | |   |  | |||     |  || | |
Db       1318 CGCGCGCGCGCCGCGGNNCGGGCGCGCGGCGGGCGCGCGGCGGGCGGGNCGCGGCGNGCG 1377

Qy        486 CGGGCCGCGGGCCCAGAGCAAGCGCG---CCGGCGACACGAGGAGGACCGCGCCCAGCAG 542
              || || ||| || | | ||  |||||    ||| | | || |  |  | ||| |      |
Db       1378 CGCGCGGCGNGCGCGGCGCGGGCGCGGGCGCGGGGNCGCGNGCGGCGCGGCGGCGGCGCG 1437

Qy        543 CACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGG 602
              | |  || || |||||||  ||  |||||  | | | |||      |      ||| |    |
Db       1438 CGCCGGCGCGCGCGCGCGGNCCNGCGCGCGCGNCGCGGCGGGCGGCGNCGCGCGCGGCCG 1497

Qy        603 CAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTAT 662
              | |          |  |  |||  ||||       || |  |      |  ||
Db       1498 CGGGCNGCGGGCGCGCGCGCGCCGGCGCGCGGCGCGGCGCGCNGCGCGGCGGGCGCGCGC 1557

Qy        663 TGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCG-GGATCTGCCGCGGATGGGCAGCCG 721
                | | |       ||  |||  ||||||| || ||  |   ||||   | || || |
Db       1558 GCGCGGCGCGCCGCGCGCGNCGGNGCGGCGGGNCGCGGCGCGCGCGCGGGCGCGGCGG 1617

Qy        722 CGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGA-TCA 780
              || |    ||  | ||      |||| |     | ||  ||  | ||||||||||    |
Db       1618 CGGGGGCGCGCGCGGCGCGCGGCGCGCGGGCGCGGNGCNGGGCNGCGCGCGCGCGNGGCG 1677

Qy        781 CCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCT 840
              | ||||| | || |||   | |||| |      |||  ||  | ||| |||    | |||
Db       1678 CGCGCGCGCAGCNGCGCCGGCGGCGCGGCGCGCGCCGGCNGGGGCGGGCGGGGNCGCGCG 1737

Qy        841 CTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCAT 900
              | |||   |       || |      ||||||  |      | |||   |  | | |
Db       1738 CCGGGCNCGGGNGACGNGCGCGCG---NGCGCGCNGGGGNCCGCGGGGCGCNGCCGNGCGG 1794

Qy        901 GTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGC 960
              |  | | |    || |||| ||| |          |||  |||  | |||  | |||
Db       1795 CGGGGNCGCGCGGCGCGNCGCGCGCGGGNNGGGNCNGGCCGCGCGGGCCGCGCGGCCGGC 1854

Qy        961 CGCGAGGCCGGG 972
               ||| | | | |
Db       1855 NGCGCGCCGGNG 1866
```

```
RESULT 6
CL515231
LOCUS       CL515231                1981 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_898_H09.v2 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_898_H09.v2, genomic survey sequence.
ACCESSION   CL515231
VERSION     CL515231.1  GI:46012551
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1981)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
```

```
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS840564; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1981
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_898_H09.v2"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match            12.5%;  Score 125.6;  DB 10;  Length 1981;
  Best Local Similarity   48.8%;  Pred. No. 3.5e-13;
  Matches  492;  Conservative    0;  Mismatches  504;  Indels   13;  Gaps    6;

Qy          1 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 60
              || | || |       | ||||| || ||||  |||  ||| |||        | |    |
Db        481 CGGCGCGGGGCGCGCGCCGGCGGCGCGCGCGGCGGGCGCGCGGCGTGCCCGCCCGCGGCC 540

Qy         61 GAGCAGGTCGGTGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCA 120
                || ||   | | ||| ||   ||||  || |||| | |  |||  |  | | ||
Db        541 CGGCCGGCGCGGGCGCGGCGGCCGGCGGCGCGCGCGGGGGGCGCGGCGGGCCCCCG-GTG 599

Qy        121 GAGGAGCGGCGTGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAG 180
              || |  ||| |||     | |||      ||   || |      |   |  |     || |
Db        600 CCGGCGGCGCGCGGCGCGCCCGGCGGCGCGCGCCGCGCGCGCGGCGCGGCCGCGCGCGCG 659

Qy        181 CCCGTGGCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCC 240
              ||  |  |||| |||||     || | ||| ||   |  |   | ||  | |  |  | |||
Db        660 CCGGGCGCGCCGCGCGCGCGCGGCCCCGGCGGCGGCCCGGCGCGCGGCGCGGGCGCGGCC 719

Qy        241 GAGCGAGAGGGCGAGCACGAGCG-GGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCG---C 296
              | |||   | |   || || ||| |||  |||||| |||||  |  |||||  | |||   |
Db        720 GCGCGTCGGCGGCCGCGCGCGCGCGGCGTCGCGCGCGCGCGTGNCGGCGCTGCGCGNTTC 779

Qy        297 GCTGCGCGGCGAGCGCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGA 356
              |   ||  ||| ||||         | | || ||||| ||  |||     | |  |  | |
Db        780 GGCNCGTTGCGCGCGCCGGTCGCGTGCGCCCGCCGCGCGGCGCGCGGGCGTNGCGCCGCG 839

Qy        357 GGACGCCCACGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTC 416
              | || |  ||  |  | ||   || || | ||| ||| ||  | |    ||| |    |
Db        840 CGGCGGCGTCGCGGCGCGCGCGCGCCGCGCGCGCGCGGCTCCGGCGCGCGCGCGTCGC 899

Qy        417 GGCGCGCG-TCGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCG 475
              ||||| ||  |||   ||| |||    ||  |  ||||   | | ||       ||  ||
Db        900 GGCGCNCGCGCGCCGACGCCGCGGCGCCGGGCGCCGCGCCGCTCGCCCGCGGCGCGCGCG 959

Qy        476 CCACCGCGAGCGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGC 535
              | | ||| || | | ||||  |  | |      ||| | ||| | |       |  |||||
Db        960 GCGCGGCGCGCCGCCNGCGGCGCGCGCGACGCAGCGACCGCGGCCCCGCCGCGCGCGCGC 1019

Qy        536 CCAGCAGCACGAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGC 595
              |   |   |   | || |||||||  ||  |||| | |    | ||   || |  |  |
Db       1020 GCGCGACGGCCTACGCGCGCGCGCGCNCCCGCGCGGCCGNGCGTCCGCGCCGCCGCCGCC 1079

Qy        596 CCATGGGCAGAGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAAT 655
              |  || |         |   | ||||  |||           || | ||   |||||  |
Db       1080 GCCCGGACCCGNCGGCCGGCGCGCCGGTGCGCGCACCGCGCCGACGGCGCCCGACGCGCT 1139

Qy        656 GCGTTATTGTGGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGG 715
              |      |  |   |       || |||      | ||  | |     ||   || | |
Db       1140 CGGCGCGCGGCGGTACCGGTCGCGCGGCGCGCGCGTCGGCGGCGGCGCGCGCGCGACGNG 1199

Qy        716 CAGCCGCGAGCCGCCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGC 775
              |   ||| ||  ||  ||| |     ||||||||||||| |||     |      |||
```

```
Db       1200 CGCGTGCGCGCGCGCGGGCCGGCGGCGTCGCGGCGCGCGCGCGCGTTCCCGGCGCGCCGC 1259

Qy        776 GAT-CACCCGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCA 834
              |     || |||  ||  ||||     | ||     | ||   |||  ||| | ||
Db       1260 GCGGNGCCGGCGGGCGCGCGCGGNACCGNCGNCGCGGGCGGCGGCGGCCGCCCCCGCCGC 1319

Qy        835 CCCGCTCTCGGCCGCGATCTTCTCGAGGTGACTGCGCGC------GTGCTCGATCACGGC 888
              |  | | || |       | || || |  |   || |||       | || ||  | ||||
Db       1320 GCGCCCCGCGCCGCACCGCGTCGCGCGCCGCACGCCCGCTCGCCGGCGCGCGCGCGCGGC 1379

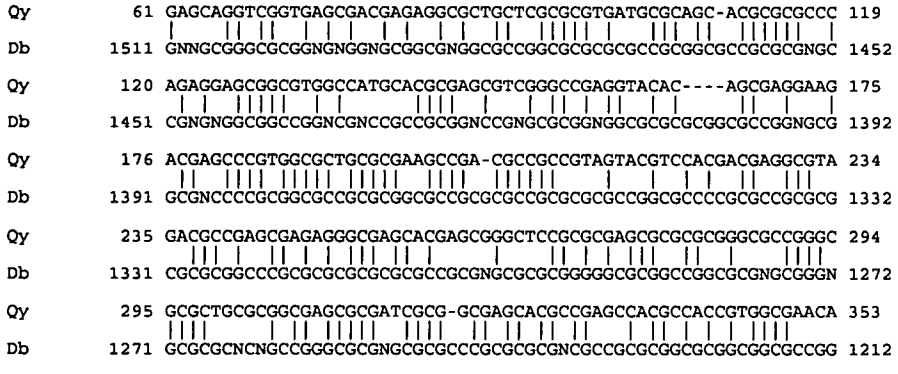
Qy        889 CTCGTTGCCCATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTC 948
              | |  |||      |   | | |  |||| || ||||| ||  | | ||   || |||
Db       1380 GTNGCCGCCGCGCCGGCCGCGGCGCGCCGCCNCGCGCGCGCGCGCCCCGGCGGCGCCCTC 1439

Qy        949 GGCGGCGTGGGCCGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCG 997
              || || || |       ||  | || | |  |  |||| | | | ||
Db       1440 CTCGACGCGGACTCGCGCGCGCGTCCGTTCGCGCGCGCCGCCGGGCGCG 1488


RESULT 7
AG171124/c
LOCUS       AG171124               1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match            12.0%;  Score 120.2;  DB 10;  Length 1798;
  Best Local Similarity   46.0%;  Pred. No. 3.4e-12;
  Matches  467;  Conservative    0;  Mismatches  529;  Indels   20;  Gaps    7;

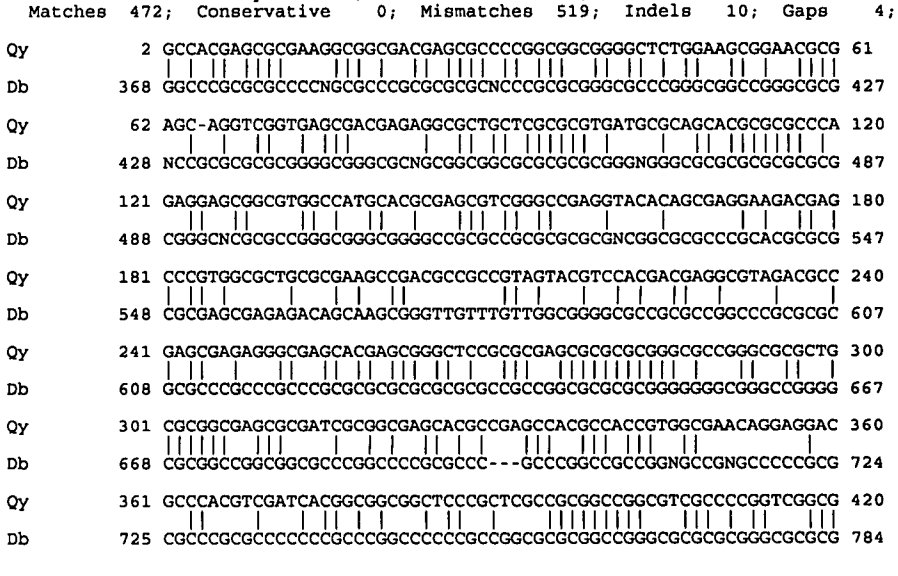
Qy          1 CGCCACGAGCGCGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGC 60
              || | ||    ||   |  | | | | | | | ||| |||  |    |       |   |
Db       1571 CGGCGCGCCGCCGCGCGGGNGNGNNGCGNGNCGCGGNCGCGNNGGCGCGCGGNNCGGNNGN 1512
```

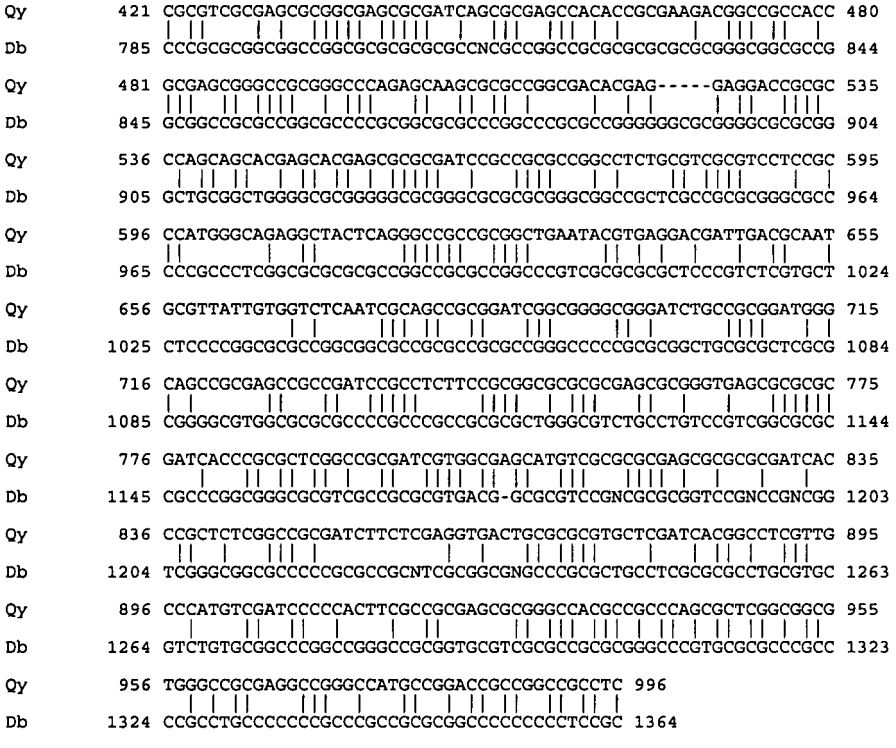


RESULT 8
BG850730
LOCUS BG850730 1409 bp mRNA linear EST 29-MAY-2001
DEFINITION 1024029B11.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG850730
VERSION BG850730.1 GI:14231914

KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 1409)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .1409
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
ORIGIN

Query Match 11.9%; Score 119.6; DB 2; Length 1409;
Best Local Similarity 47.2%; Pred. No. 4.5e-12;
Matches 472; Conservative 0; Mismatches 519; Indels 10; Gaps 4;





RESULT 9
AG032943/c
LOCUS AG032943 1542 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-007N10.F, genomic survey sequence.
ACCESSION AG032943
VERSION AG032943.1 GI:16559816
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1542)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: -21M13

LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .1542
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-007N10.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

Query Match 11.9%; Score 119.6; DB 10; Length 1542;
Best Local Similarity 45.5%; Pred. No. 4.5e-12;
Matches 456; Conservative 0; Mismatches 540; Indels 6; Gaps 4;

```
Db        666 CGCGGCCGGCGCGCCGGGGGCGCGGCGGCCCNCCGCGCCCCCGCCGCCGCGNCCGCGTGN 607

Qy        836 CCGCTCTCGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTG 895

Db        606 GCGCGGCGCCGCGCGGCGCCCGCGCCGGCGCCGGCNNGGGCCGCGCGGGCCGCGGGGGGC 547

Qy        896 CCCATGTCGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCG 955

Db        546 NCGCGGCGGGCGCGCGGCNNGCCGCCGCGCGGGGCNCNGCCCGCCCGCGCCCNCCCNNCC 487

Qy        956 TGGGCCGCGAGGCCGGGCCATGCCGGACCGCCGGCCGCCTCG 997

Db        486 CCCGCCGCCGGCCGGCNCNNNGGCGGCGCGGCCGCGGGGCCG 445


RESULT 10
BG809984/c
LOCUS       BG809984                1516 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  mgct002xd11f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct002xd11f 5', mRNA sequence.
ACCESSION   BG809984
VERSION     BG809984.1  GI:14180964
KEYWORDS    EST.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES             Location/Qualifiers
     source          1. .1516
                     /organism="Magnaporthe grisea"
                     /mol_type="mRNA"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct002xd11f"
                     /dev_stage="Germinated conidia on apprressorium-inductive
                     surface"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed from conidia germinated for 5-8 hr
                     on an inductive surface. The library contains over 55,000
                     clones with average insert size of 1.5 kbp."
ORIGIN

  Query Match           11.9%;  Score 118.8;  DB 2;  Length 1516;
  Best Local Similarity   45.4%;  Pred. No. 6.3e-12;
  Matches  444;  Conservative    0;  Mismatches  520;  Indels   15;  Gaps    3;

Qy         29 GCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGGTGAGCGACGAGAGGCGC 88

Db       1208 GCTCCGCGCGCGCGGCGNCGGGCGCGCGCGTGCGGCGGCGGCGNGNGNNNCGCGCGCGGG 1149

Qy         89 TGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTGGCCATGCACGCGAGC 148

Db       1148 CNCGCNCGNGGGCGCGCGGCNGCGNGCGCGCGGCGGGCNCCCGCGCNCGCGGGGCGCGCC 1089

Qy        149 GTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGCGCGAAGCCGACGCCG 208

Db       1088 NNCNCGGCGCCGCGCGCCGGGCGGCCGCGGCGCGCGGCGGGCCGCGGGCGTCCGGNGC 1029
```

```
Qy        209 CCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCACGAGCGGGCTC 268

Db       1028 GGCTCGCCGCGGCGCGNGGCGCGGGCCGCGCGGCGCGGCGCCGCGGCCGCCCGCGCCGCC 969

Qy        269 CGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGATCGCGGCGAGCACG 328

Db        968 GGCGGGGGGGCGCGCGGGGCCGGGGCGGCCCGGCGCGGGGCGGCCGCGGCNGCGNGCCGC 909

Qy        329 CCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCACGGCGGCGGCTCCC 388

Db        908 GGCCGCGCCGCGGCGCGCGCGCCCCGCGCCCNGCCCGCGCGGCGCGCGCCGNCCGCCCCC 849

Qy        389 GCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGCGCGGCGAGCGCGATCA 448

Db        848 CGCCCCGCGCGCGCGCCTGGCGCGCNNCGCGGCGGCGCGCGCCGGCGCGCNCGCCGCGGC 789

Qy        449 GCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGC----------GGGCC 498

Db        788 GGCCNGCCGCCGCGCCCGGCNCCGCCGGCCGCGCGNGGGGCCCCCNNCCGCGNCGGCGCC 729

Qy        499 CAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCACGAGCGCG 558

Db        728 CCGCCCCGCCGCGCGCCCGCCGGCCGCGCGCGCNCNCGCCGGCCCGCGCCGGCGGG 669

Qy        559 CGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGAGGCTACTCAGGG 618

Db        668 CGCGGCGCGGGCGCCGGGCGCGNGGCG-GGCGGGCGGCGCNGGGCGGCGCCGGCGCCGCG 610

Qy        619 CCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGTCTCAATCGCA 678

Db        609 CGGCCCCCGCGGCGCCCGGCCGCGNGCCCGCCCCGCNGCGCGCGCCGCGCGCGGCNGCCC 550

Qy        679 GCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCGCCGATCCGCC 738

Db        549 GCGGCGGCGCCGCGCGGCGNGCNGGNGNCCGCCCCCCCCCCGCGGCCGCNCCCCNNCNCN 490

Qy        739 TCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCTCGGCCGCGAT 798

Db        489 CCGGCGCCCGCGGGCGCGCGCGCGCGCGCGCGCGCCGCNCGCCGCGCGGNCGGCGCGGCG 430

Qy        799 CGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCGCGATCTTCTC 858

Db        429 CGCGCCGGCCGCNGCCGCGCGCGGCCCGCGCCGNGGGNGCGNCGGGNCCGGCGGCGGCGC 370

Qy        859 GAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGATCCCCCACTTCGC 918

Db        369 GCGCCCCGCGCGCGCGGGCCCG----GGGCCGCGCGGCCGCGCGGGGCGGCCGGCGNCGC 314

Qy        919 CGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGAGGCCGGGCCATGC 978

Db        313 GCCGCGGGCCGGCCCGCCCGCGCCGCGCGGGGCGCCGCGGGCGGNCCGCGCGCCCGGCGC 254

Qy        979 CGGACCGCCGGCCGCCTCG 997

Db        253 CGGGCGGCGCGCGCGCCCG 235


RESULT 11
AG032979
LOCUS       AG032979                1452 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-007004.F, genomic survey sequence.
ACCESSION   AG032979
VERSION     AG032979.1  GI:16559852
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1452)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1452
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-007004.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             11.8%;  Score 118.6;  DB 10;  Length 1452;
  Best Local Similarity   50.4%;  Pred. No. 6.9e-12;
  Matches  406;  Conservative    0;  Mismatches  388;  Indels   12;  Gaps    6;

Qy        187 GCGCTGCGCGAAGCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGA 246
              |||| |||||    || ||||  ||       |    || || || |  |  | || | |
Db        499 GCGCCGCGCGCGCGCGCCGCCCGCGCGNNGGGNNGGCGGCGCGGNGCCGCGGGCGCGNGC 558

Qy        247 GAGGGCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGC 306
              | |         | |||  | |   | |   ||||||| |||   ||||    |||||
Db        559 GCGCCGCGCGNNGCGCGCCCGCGCGGGGCCGGCGCGCGCGCGGGCGGCGGCGGGCGCGCG 618

Qy        307 GAGCGCG-ATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCA 365
                 |||   ||| ||  || || ||  |  ||    || | ||     | |  |||
Db        619 NNNGGCGCCGCGCCGCCGGCCCGGCGGCCGGCGGGNGCGNGNCGNCGGCGCGNCCGCGGG 678

Qy        366 CGTCGATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGT 425
              ||  |      ||| ||   || | ||||   ||| |  ||| || || | | ||  ||
Db        679 CGCGGCCGCGCGCGCCGCGCGCCCCGCGCCCGCGCCCNCCGCGCGCCCGCCCGCCGGCCC 738

Qy        426 CGCGAGCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAG 485
              |||   | |  |  || ||  |  |||   || | |||||  | ||  |||  || || |
Db        739 CGCCCCCCCCCCCCGCCCGCCCCCCGC---CCGCGCCGCGCGGCCGCGCGCGCCCNCGCG 795

Qy        486 CGGGCCGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCAC 545
              || || ||| || |    |  ||||||||      | |    |||||  |   | || |
Db        796 CGCGCGGCGCGCGCGCGCCGCGCGCGCCGCGCCGCCCCGCCCCGCCGCGGCGCCCCGCGC 855

Qy        546 GAGCACGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAG 605
              |    || || |||    ||  ||| |||  | |  || |||  ||  ||| |  | ||
Db        856 GCCGCCGCGCCCGCCGGCCCGCGCCCCGCGCGCGCCGCCGCCGCCGGCGCGCGCGCGCGC 915

Qy        606 AGGCTACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGT 665
               |          || ||||| ||       || |  |  |      | |   ||      |
Db        916 GGCGCCGCGGCCGCGGCCGCCGCGCCGCGCGCGCCGCGGCGCCGCCCGCGGCCGGGCCGC 975

Qy        666 GGTCTCAATCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAG 725
              |  | |   |||   ||||      | |  |  |  |   | |||  | || ||||||  |
Db        976 GCGCGCG-CCGCGCGCGCGCGCCCCGCCCCNGNGCGCGCCCGGGCGCGCCGCCGCGCG 1034

Qy        726 CCGCCGATCCGC---CTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACC 782
               |||      |||    |     ||||| | ||| || || || |   |||||||       |
Db       1035 GCGCGCGCGCGCGCGCGGCCCCGCGCCCGCCCGCGCCCGCGGCCGCGCGCGNCGCGCCG 1094

Qy        783 CGCGCTCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCT 842
              | ||| |||||||   || | || ||  |  || |||| ||   ||||  |   |||
Db       1095 CCCGCGCGGCCGCNGCCGCGCCGCGCCGGGGGC-CGCGGGCCGCCGCGGNCCGGCGCCGC 1153

Qy        843 CGGCCGCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGT 902
                ||   |  | | || |    | ||||||  || ||  |   ||| ||  ||||
Db       1154 GCGCGCGGCGCGNCGCGGGCCCGCCGCGCGCCCGCNCGCGC---GCCGCGCNGCCCGGCG 1210

Qy        903 CGATCCCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCG 962
              |   | | |    |||||| || || ||    | || |  ||| ||  |||| | ||||
Db       1211 CNCGCGCGCCNCGCGCCGCGCGCCCGCGCGCGCCGGCNCGCCGCGCGCGGGCGCGCGCCG 1270

Qy        963 CGAGGCCGGGCCATGCCGGACCGCCG 988
              ||       | ||  ||||  |||||
Db       1271 CGCCCGGCGCCCCGGCCGCCGCGCCG 1296


RESULT 12
CL470317
LOCUS       CL470317               1280 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_140_A11.v1 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_140_A11.v1, genomic survey sequence.
ACCESSION   CL470317
VERSION     CL470317.1  GI:45936028
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1280)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS806754; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1280
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_140_A11.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             11.8%;  Score 118.2;  DB 10;  Length 1280;
  Best Local Similarity   48.6%;  Pred. No. 8.3e-12;
  Matches  390;  Conservative    0;  Mismatches  409;  Indels    3;  Gaps    3;

Qy        199 GCCGACGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCAC 258
              || |  ||  ||   |  |  |  |  | || |  | ||| | | | | ||| |    |
Db        213 GCGGGGGCGCCCCGCGCGGGGGCGGGGGGGGGGGGCGGCGCGGCGGGCGGGGGGGGCGGC 272

Qy        259 GAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGATCGC 318
              | ||| ||   |   | |||||||||| |  ||| | | || ||| |  |          ||
Db        273 GGGCGCGCGGGGGCGGCGCGCGCGCGGCCCGCGTGGGCGGCGGGCGGNCGCGCGGGGC 332
```

```
Qy        319 GGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCACGGC 378
Db        333 GGCGGCGGCGGGGGGCCGGGCGGCGGCGNCGGCCCGCGCGGNCGGGCCGGCGGGGGGGGC 392

Qy        379 GGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGC-GCGGC 437
Db        393 GGCGGGGCGGGCGGGCGGCGCGCGCGCGGGGGCGCGCGGGCGGGCGGCGGGGCCGGGGGC 452

Qy        438 GAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGCGGGC 497
Db        453 GGGCGCGGGCGGGGCGCGCGGGCCGGCGGNGGCGGCGGCGGCGCGGGGCGGGCGCCGGGG 512

Qy        498 CCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCACGAGCGC 557
Db        513 CGCGCGCGGGCGCGCGGGGGGGGGGCGGGGGCGGGCGGGCGGGTGCGCGGGCGGCGGCGC 572

Qy        558 GCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGC-CCATGGGCAGAGGCTACTCAG 616
Db        573 NCGCGCGCGCGCGGCGTGCGCGAGGCCGCGGCCCGCGCGCCGCGCGCGGCGGGCGGGGCG 632

Qy        617 GGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGTCTCAATCG 676
Db        633 CGGCGCCGCGCGCGCGTTCCCGCGGCCGCGGGCGCGCGCCGCCGCCGCCGCGGCGCGGCG 692

Qy        677 CAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCGCCGATCCG 736
Db        693 GGGCGTCGGGCGGGCGGGCGGGGGCGGCGCGGGCGGGCGCGGGCGCCCGCGGCGGCCGCG 752

Qy        737 CCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCTCGGCCGCG 796
Db        753 GCGGCGCCGTGGCGGGCCGGCGCGCGCGGCGGGCGCGGCGGAGGCGCGCGGCGGGGGCCG 812

Qy        797 ATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCGCGATCTTC 856
Db        813 CGGCGGGGCGGCGCGGCGCGGGCGCGNACCNCGCGCGCGCCGGAGCCCGCGCCGGCGCGC 872

Qy        857 TCGAGGTGACTGCGCGCGTGCTCGATC-ACGGCCTCGTTGCCCATGTCGATCCCCCACTT 915
Db        873 CGGCGNCCGCGCCGCGCCGCCGCGCGCGGCGGCCGCGTNCCGCCCGCGGCGACGGGGGCC 932

Qy        916 CGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGAGGCCGGGCCA 975
Db        933 GCGCGNCGGCGCGCGGCGGGCGCGCGGCGCGGCGGCGGGCGCGGGCGCGNTGCGGCGCGG 992

Qy        976 TGCCGGACCGCCGGCCGCCTCG 997
Db        993 GACGCGCGCGCGCGCCGCCGCG 1014
```

```
RESULT 13
AG064051
LOCUS       AG064051               1126 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-053A06.F, genomic survey sequence.
ACCESSION   AG064051
VERSION     AG064051.1  GI:16615853
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1126)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1126
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-053A06.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
```

```
  Query Match            11.7%;  Score 117.6;  DB 10;  Length 1126;
  Best Local Similarity  48.7%;  Pred. No. 1.1e-11;
  Matches  442;  Conservative    0;  Mismatches  460;  Indels    5;  Gaps    5;

Qy         85 GCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCGTGGCCATGCACGC 144
Db        127 GCGNNGCGCCCCCGCCCGGCGCGGCGGCCGCGGNGGGGGGCCCGGGGCCGCCGGNGCCGC 186

Qy        145 GAGCG-TCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCTGCGCGAAGCCGA 203
Db        187 GGGCGCCCGCGGCGGCGGCGCCGGGGCGGGGGCGGCGCGCGCCGCGGGGGNGGGGGCGCC 246

Qy        204 CGCCGCCGTAGTACGTCCACGACGAGGCGTAGACGCCGAGCGAGAGGGCGAGCACGAGCG 263
Db        247 GGCGGGNCGGNGGNGNGGCCGCCGGCGCCCGCCCGGCCCGCGGCGCGGGCGGCCGCGCGG 306

Qy        264 GGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGATCGCGGCGA 323
Db        307 CGCGCCGGGCGGGCGCGCGCGCGGGGCGCGGGGCGCGCGCGG-GCGCCCGCGGCCCGCGG 365

Qy        324 GCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCACGGCGGCGG 383
Db        366 CGGCGGNGGCCGCCGGCGGGGGGCGCGGGCGCCGGGNGCCCCCGGGCGGCCCGGGCGGCG 425

Qy        384 CTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGCGCGGCGAGCGC 443
Db        426 CGCGGGCGGGCGGGGGCGCGCGGCGCGCGCGCCGGGGCGGGGCGCGCGCGGCGCGCGGCG 485

Qy        444 GATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGCGGGCCCAGAG 503
Db        486 GGCGCGGGCGCGCGCCCCCCGCGGCGGGCGCGGCGCGCGGCGGGCCGGCGCGCCGGCGG 545

Qy        504 CAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCACGAGCGCGCGATC 563
Db        546 CCGCCGCGGCGCGCGCGCCCGCGCGGCGCCACGACGCCGCGCGCGCGCGCGGCGGCGCGC 605

Qy        564 CGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGAGGCTACTCAGGGCCGCC 623
Db        606 GGGCGC-CGGGGAGGGGCCGGCCGCGTCCCGCCGCGGCGCCCGCGCGCCCCGGGGCCGCC 664

Qy        624 GCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGTCTCAATCGCAGCCGC 683
Db        665 GCCGGAGAGCCGCGGGGGGGCGGCGCGCGGGGCCGCGCACGCGCGCCGGACCGGGCGCGC 724

Qy        684 GGATCGGCGGGGCGGGATC-TGCCGCGGATGGGCAGCCGCGAGCCGCCGATCCGCCTCTT 742
Db        725 GCCCCGCGCGCGCGCGCGCGGCGCGGCGGCGCGCCGCGCAACCCCCGCGCGGCGCGCG 784

Qy        743 CCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCTCGGCCGCGATCGTG 802
```

```
Db        785 GCGGCGGCCGCGGCGGCGCCGGCGCGCGCGCGGGGGCGCGCGGGCGCGAGCCCGGCGGCC 844

Qy        803 GCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCGCGATCTTCTCGAGG 862

Db        845 GCGCGCCCCGCGGCCGCGGCCCCCCCCCCGCCGGGGACCGCGCCCCGGGGCGCGTCGGCC 904

Qy        863 TGACTGCGCGCGTGCT-CGATCACGGCCTCGTTGCCCATGTCGATCCCCCACTTCGCCGC 921

Db        905 GGACGAGGGGCGCGGGCCGGCCGCGGGGGCGCGGCGCGGGCCGCGCGGGCGGCGCGCCGC 964

Qy        922 GAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGAGGCCGGGCCATGCCGG 981

Db        965 CCCGGCCGCGCCGCGCCGGGCGGGCGCCGGGCGACGGGCCGCACGGCGGGGCGCGCCGGG 1024

Qy        982 ACCGCCG 988

Db       1025 GCCGCCG 1031
```

```
RESULT 14
AG039481/c
LOCUS       AG039481               1299 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-016J15.F, genomic survey sequence.
ACCESSION   AG039481
VERSION     AG039481.1  GI:16568206
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1299)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1299
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-016J15.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             11.7%;  Score 117.6;  DB 10;  Length 1299;
  Best Local Similarity   49.2%;  Pred. No. 1.1e-11;
  Matches  366;  Conservative    0;  Mismatches  367;  Indels   11;  Gaps    4;

Qy        255 GCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGCGCGA 314

Db       1225 GCGCGGCGGGNNGNCGCGCGCGCGCGCNNCCCGNGCGCCCCGCGCGCGCGCCGCGCGNGG 1166
```

```
Qy        315 TCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCGATCA 374

Db       1165 GCGCGCGGGNCGGGCGGCCGNCCGCCGCGCGGGCCGCGCGCGCGCGNGGGGCGCGGCCGC 1106

Qy        375 CGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGAGCGC 434

Db       1105 GCGGCGCGCGGCGCGGGCGCCGCGCCGCGCGGCGCGCGCGCGNCGCGCGCGCGCGCGCGC 1046

Qy        435 GGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGCCGCG 494

Db       1045 GGNG-GCCCGCGCGGCGCGCGGCGCGGCGCCGGCGCGCGCGGCGCGNCNCCCGCCGCG 987

Qy        495 -GGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCAGCAGCACGAGCACGA 553

Db        986 CCGCGCCGCCCGCGCGCGNCCGCCGCGGNNCGGCGGGCGCGCCGCGNNGGCCGGCGGNCG 927

Qy        554 GCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGAGGCTACT 613

Db        926 GCGCGGCNCGNCGCGCNGCGCGCNGCGCGCGGCGCGCCGCGCGCGCGCGCCGCGGCGCGC 867

Qy        614 CAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGTCTCAA 673

Db        866 GNGCGCGGGCGCGCGGCGGCGCGCGCGCGCCGNGCGCGGGGGGCGCGCGCGCGGCCGCCN 807

Qy        674 TCGCAGCCGCGGATCGGCGGGGCGGGATCTGCCGCGGATGGGCAGCCGCGAGCCGCCGAT 733

Db        806 CCGCGCGCGCNCNCGCGCGNCGCGGCGNNCGGCGCGGCGCGGCGCGCGGCGCGCCGGCGG 747

Qy        734 CCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCACCCGCGCTCGGCC 793

Db        746 CGGCGGCGCGCGCGCGCCGCGCGCGCGGCGGGGCNCGCGCGCG-CGGCGCGCGCGCGGGC 688

Qy        794 GCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCCGCGATC 853

Db        687 GCGGCCGCGGCGGGCGCCGCGCGNGGGGCGGCGCGNGGGCGGCGGGGGCCGGNCGCGCGN 628

Qy        854 TTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGATCCCCCAC 913

Db        627 GGGGCGGGGGGCGCGCGGGCGGCGGCGCGCGGCGGGGCGCGG--------GGGCGCGCGC 576

Qy        914 TTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGCTCGGCGGCGTGGGCCGCGAGGCCGGGC 973

Db        575 GGCGCGGCGCGCGCCGGGCGCGCCGCGCCGGCCGCCGCCGGCGCCGCCGCGCCGCCGGCC 516

Qy        974 CATGCCGGACCGCCGGCCGCCTCG 997

Db        515 CGGCGCGCCNCGCGCGGCGCGCCG 492
```

```
RESULT 15
AG032979/c
LOCUS       AG032979               1452 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-007O04.F, genomic survey sequence.
ACCESSION   AG032979
VERSION     AG032979.1  GI:16559852
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1452)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
```

```
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1452
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-007004.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             11.7%;  Score 117.2;  DB 10;  Length 1452;
  Best Local Similarity   45.9%;  Pred. No. 1.2e-11;
  Matches  431;  Conservative    0;  Mismatches  500;  Indels    8;  Gaps    4;

Qy         12 CGAAGGCGGCGACGAGCGCCCCGGCGGCGGGGCTCTGGAAGCGGAACGCGAGCAGGTCGG 71
              ||  || |  |  | ||| | |||||  | |||   ||   |||     | ||  | | |
Db       1317 CGCNGGNGATGGNGCGCGNCGCGGCGCGGCGGCCGGGGCGCCGGGCGCGGCGCGCGCCCG 1258

Qy         72 TGAGCGACGAGAGGCGCTGCTCGCGCGTGATGCGCAGCACGCGCGCCCAGAGGAGCGGCG 131
              | ||| || |  |    || || ||| |  ||||    ||||||  |   ||   |
Db       1257 CGCGCGGCGNGCCGGCGCGCGCGGGCGCGCGGCGCGNGGCGCGCGNGCGCCGGGCNGCGC 1198

Qy        132 TGGCCATGCACGCGAGCGTCGGGCCGAGGTACACAGCGAGGAAGACGAGCCCGTGGCGCT 191
              |  |  |  || | |||  | |      |       ||   |    ||   |||   |||
Db       1197 GGCGCGCGNGCGGGCGCGCGGCGGGCCCGCGNCGCGCCGCGCGCGCGGCGCCGGNCCGCG 1138

Qy        192 GCGCGAAGCCGACGCCGCCGTAGTACGTCCACGA-CGAGGCGTAGACGCCGAGCGAGAGG 250
              |||    || | | ||| ||  |  || |  ||  || |  |  | |||   ||| | ||
Db       1137 GCGGCCCGCGGCCCCCGGCGCGGCGCGGCNGCGGCCGCGCGGGCGGCGCGNCGCGCGCGG 1078

Qy        251 GCGAGCACGAGCGGGCTCCGCGCGAGCGCGCGCGGGCGCCGGGCGCGCTGCGCGGCGAGC 310
              || |  || | | |   |||| |  |||||||| ||||  | ||||| |||  |||  |
Db       1077 CCGCGGGCGCGGGCGGGGCGCGGGGCCGCGCGCGCGCGCGCGCGCGGCGGCGCGCCC 1018

Qy        311 GCGATCGCGGCGAGCACGCCGAGCCACGCCACCGTGGCGAACAGGAGGACGCCCACGTCG 370
              | |  |||   |     | || | | |||   || |  |  |  | |    ||  |  ||
Db       1017 GGGCGCGCNCNG----GGGCGGGGCGCGCGCGCGCGCGGCGCGCGCGGCCCGGCCGCG 962

Qy        371 ATCACGGCGGCGGCTCCCGCTCGCCGCGGCCGGCGTCGCCCCGGTCGGCGCGCGTCGCGA 430
                |   || |||||  | |||    ||  | | ||| |  |   | ||        ||||
Db       961 GGCGGCGCCGCGGCGCGCGCGGCGCGGCGGCCGCGGCCGCGGCGCCGCGCGCGCGCGCGC 902

Qy        431 GCGCGGCGAGCGCGATCAGCGCGAGCCACACCGCGAAGACGGCCGCCACCGCGAGCGGGC 490
              ||||||   |||  |   | | ||    | |||    |||| ||   || |     ||
Db       901 CGGCGGCGGCGGCGCGCGCGGGGCGCGGGCCGGCGGGCGCGGCGGCGCGCGGGGCGCCGC 842

Qy        491 CGCGGGCCCAGAGCAAGCGCGCCGGCGACACGAGGAGGACCGCGCCCCAGCAGCACGAGCA 550
              |||||  |  |    |||   ||          | |  ||||||  |    |  || ||
Db       841 GGCGGGGCGGGGCGGCGCGGCGCGCGCGGCGCGCGCGCCGCGCGCGCGCGNGGGCGCGCG 782

Qy        551 CGAGCGCGCGATCCGCCGCGCCGGCCTCTGCGTCGCGTCCTCCGCCCATGGGCAGAGGCT 610
              ||  ||||||   ||     |  |||    | |  | |    | |  |  | |    |||
Db       781 CGGCCGCGCGGCGCGGGCGGGGGGCGGGCGGGGGGGGGGGGCGGGGCCGGCGGGCGGGCG 722

Qy        611 ACTCAGGGCCGCCGCGGCTGAATACGTGAGGACGATTGACGCAATGCGTTATTGTGGTCT 670
                  || ||| |  || |    || |  |      | |||
Db       721 CGCGGNGGGCGCGGGCGCGGGGCGCGCGGCGCGCGCGGCCGCGCCCGCGGNCGCGCCGNC 662

Qy        671 CAATCGCAGCCGCGGATCGGCGGGGCG--GGATCTGCCGCGGATGGGCAGCCGCGAGCCG 728
                |||  |||| |  || |||| ||  ||  | ||  |    | ||  ||||   |
Db       661 GNCNCGCNCCCGCCGGCCGCCGGGCCGGCGGCGCGGCGCCNNNCGCGCGCCCGCCGCCGC 602

Qy        729 CCGATCCGCCTCTTCCGCGGCGCGCGCGAGCGCGGGTGAGCGCGCGCGATCAC-CCGCGC 787
              |||  |   | |   | | ||||| ||| ||   |    ||||||  ||  | | | || |
Db       601 CCGCGCGCGCGCCGGCCCCGCGCGGGCGCGCNNCGCGCGGCGCGCNCGCGCCCGCGGCNC 542

Qy        788 TCGGCCGCGATCGTGGCGAGCATGTCGCGCGCGAGCGCGCGCGATCACCCGCTCTCGGCC 847
               |||||  |    || ||  | |||||| ||||      |  |    | ||||
Db       541 CGCGCCGCCNNCCCNNCGCGCGGGCGGCGCGCGCGCGCGGCGCGCCNNCGCGGCGCGGCG 482

Qy        848 GCGATCTTCTCGAGGTGACTGCGCGCGTGCTCGATCACGGCCTCGTTGCCCATGTCGATC 907
              |      |              ||   ||   |  |      || || | |   ||   |
Db       481 CCNCCNNNCCNNNTANATAAATGCNGACGCCGCACAATATATTCTTTTCTCGGCTCNNNC 422

Qy        908 CCCCACTTCGCCGCGAGCGCGGGCCACGCCGCCCAGCGC 946
              ||     | |  |      ||  |  | ||  |   |||
Db       421 NCCNNNNGCNCGNCCCNNCCGCNCACCACCCACACACGC 383


Search completed: April 28, 2006, 12:17:45
Job time : 4963.09 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 5541.08 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_55000_55999
Perfect score: 1000
Sequence: 1 tcacggcgatggtctgcgat.....cccccgatgcagctcccgcg 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	84868	1	DQ013294	DQ013294 Polyangiu
	2	121.4	12.1	23451	1	AB070952	AB070952 Streptomy
c	3	121.4	12.1	110000	1	BA000030_39	Continuation (40 o
	4	120.2	12.0	4296	6	AR621890	AR621890 Sequence
c	5	120.2	12.0	17727	6	AR619757	AR619757 Sequence
c	6	112.4	11.2	110000	1	BA000030_72	Continuation (73 o
	7	110	11.0	110000	1	CP000010_11	Continuation (12 o
	8	108.2	10.8	22449	1	SPSNBCDE	Y11548 S.pristinae
	9	108.2	10.8	22449	1	SPSNBCGEN	X98690 S.pristinae
	10	107.8	10.8	127559	1	AY787762	AY787762 Streptomy
c	11	107.6	10.8	71138	6	CQ800553	CQ800553 Sequence
c	12	107.6	10.8	71138	6	CQ813424	CQ813424 Sequence
c	13	107.6	10.8	89153	1	NSP561198	AJ561198 Actinomad
	14	106.4	10.6	7110	6	AR636660	AR636660 Sequence
	15	106.4	10.6	7110	6	AX097461	AX097461 Sequence
	16	106.4	10.6	14759	1	AE004755	AE004755 Pseudomon
	17	106	10.6	110000	1	AP006618_12	Continuation (13 o
	18	103.6	10.4	1094	10	PM7G11B	AL685196 Penicilli
	19	102.8	10.3	5862	6	CS073958	CS073958 Sequence
	20	102.8	10.3	5862	6	AX598607	AX598607 Sequence
	21	102.8	10.3	52101	6	CS073944	CS073944 Sequence
	22	102.8	10.3	52101	6	AX598593	AX598593 Sequence
	23	102.6	10.3	276800	1	SCO939115	AL939115 Streptomy
c	24	102	10.2	1393	10	PM11H12G	AL684264 Penicilli
	25	101.8	10.2	1007	10	PM3H11G	AL685749 Penicilli
	26	101.6	10.2	976	10	PM12D6G	AL684370 Penicilli
	27	101.6	10.2	110000	1	BX571965_20	Continuation (21 o
	28	101.4	10.1	79023	14	AC021951	AC021951 Homo sapi
	29	101.2	10.1	80133	14	AC021347	AC021347 Homo sapi
	30	101.2	10.1	219952	14	AC084804	AC084804 Mus muscu
c	31	100.8	10.1	110000	1	BX571966_21	Continuation (22 o
c	32	100.8	10.1	110000	1	BX571966_22	Continuation (23 o
c	33	100.8	10.1	110000	1	CP000011_17	Continuation (18 o
	34	100	10.0	166401	14	CR352260	CR352260 Danio rer
	35	99.8	10.0	1794	13	HS1ICPL	M12240 Herpes simp
	36	99.2	9.9	159997	14	CR352334	CR352334 Danio rer
c	37	99	9.9	303091	14	AC084799	AC084799 Mus muscu
	38	98.8	9.9	55972	1	AF386507	AF386507 Streptomy
	39	98.6	9.9	24459	6	AR623603	AR623603 Sequence
c	40	98.6	9.9	32241	6	AR619852	AR619852 Sequence
	41	98.6	9.9	66669	1	AME16952	Y16952 Amycolatops
c	42	98.2	9.8	205691	14	AC087227	AC087227 Mus muscu
	43	98	9.8	1002	6	AR622138	AR622138 Sequence
	44	98	9.8	1008	6	AR618722	AR618722 Sequence
	45	98	9.8	6297	6	CS073980	CS073980 Sequence

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRAREAEREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPFFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

```
                GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
     CDS        complement(1992. .3440)
                /note="ORFu2; similar to Pseudomonas putida KT2440 PP4696;
                putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32957.1"
                /db_xref="GI:63099934"
                /translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
                APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
                QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
                KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
                EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
                VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
                ELTETEIARRLGISRKALWERRQRLGIPRPRS"
     CDS        3587. .5548
                /note="ORFu3; ORFu3p; similar to Leptospira interrogans
                AAN50474"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32958.1"
                /db_xref="GI:63099935"
                /translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
                VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
                LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
                YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
                SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
                IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
                VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
                AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
                GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
                AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
                SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
                VVAGFGATGPLRVMETLPEQYLFLTRVRP"
     CDS        complement(5702. .7231)
                /note="ORFu4; ORFu4p; similar to Burkholderia fungorum
                hypothetical protein in ZP_00033665"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32959.1"
                /db_xref="GI:63099936"
                /translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
                AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
                PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
                VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
                MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
                QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPFSANQRPASTLAQWRVGLEHP
                GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
                GRVITVHFHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
                ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
                P"
     CDS        complement(7561. .8931)
                /note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440
                PP4696; putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32960.1"
                /db_xref="GI:63099937"
                /translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
                APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
                AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
                VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
                VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
                PPPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
                IPRPRS"
     CDS        complement(8928. .11939)
                /note="ORF1; ORF1p; similar to Pseudomonas putida KT2440
                HisK (PP4695) in GenBank Accession Number AAN70268"
                /codon_start=1
                /transl_table=11
                /product="putative sensory box histidine kinase"
                /protein_id="AAY32961.1"
                /db_xref="GI:63099938"
                /translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
                YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
                AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
                FTAFFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
                HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
                FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
                VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
                LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
                IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPAREEREAAR
                LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
                LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
                RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
                LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
                GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
                DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
                DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
                IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
                LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
     CDS        complement(12405. .13379)
                /note="ORF2; ORF2p; similar to family of known or putative
                phosphotransferases, including macrolide
                2-phosphotransferases; similar to Bacillus subtilis YcbJ,
                Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273
                MphB and Staphylococcus aureus plasmid pSR1 MphBM"
                /codon_start=1
                /transl_table=11
                /product="putative macrolide kinase"
                /protein_id="AAY32962.1"
                /db_xref="GI:63099939"
                /translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
                DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
                VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
                IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
                DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
                AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
     CDS        13368. .14663
                /note="ORF3; ORF3p"
                /codon_start=1
                /transl_table=11
                /product="putative Ser/Thr protein kinase"
                /protein_id="AAY32963.1"
                /db_xref="GI:63099940"
                /translation="MFGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
                ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
                RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
                GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
                AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
                DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
                TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
                LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
     gene       15731. .33766
                /gene="dszA"
     CDS        15731. .33766
                /gene="dszA"
                /note="PKS (polyketide synthase) subunit modules 1-4a"
                /codon_start=1
                /transl_table=11
                /product="DszA"
                /protein_id="AAY32964.1"
                /db_xref="GI:63099941"
                /translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
                PRAHYSPDFNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
                IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT
```

```
                    LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                    LSPDGLCKTFDADANGYVPGDGVGVLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                    ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                    RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                    VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                    RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                    DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                    RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                    RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                    ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                    RAQFTFKKFLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                    LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                    PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                    PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                    EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE

  Query Match             100.0%;  Score 1000;  DB 1;  Length 84868;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-74;
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCACGGCGATGGTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62574 TCACGGCGATGGTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGG 62633

Qy         61 CGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62634 CGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGC 62693

Qy        121 AGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62694 AGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGC 62753

Qy        181 TCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62754 TCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCG 62813

Qy        241 CGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62814 CGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGG 62873

Qy        301 CGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62874 CGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCT 62933

Qy        361 CCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62934 CCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGG 62993

Qy        421 CGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62994 CGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCG 63053

Qy        481 AGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63054 AGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCT 63113

Qy        541 CGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63114 CGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGC 63173

Qy        601 TCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63174 TCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGC 63233

Qy        661 CCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63234 CCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTC 63293

Qy        721 CGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63294 CGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGG 63353

Qy        781 CCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63354 CCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATC 63413

Qy        841 CCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63414 CCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCC 63473

Qy        901 TGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63474 TGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCG 63533

Qy        961 ACGACCTCGCCGGGCTCCCCCCCCCCGATGCAGCTCCCGCG 1000
              |||||||||||||||||||||||||||||||||||||||||
Db      63534 ACGACCTCGCCGGGCTCCCCCCCCCCGATGCAGCTCCCGCG 63573


RESULT 2
AB070952
LOCUS       AB070952               23451 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis peptide-3 biosynthetic gene cluster.
ACCESSION   AB070952
VERSION     AB070952.1  GI:15824182
KEYWORDS    .
SOURCE      Streptomyces avermitilis
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
   PUBMED   11572948
REFERENCE   2  (bases 1 to 23451)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .23451
                     /organism="Streptomyces avermitilis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33903"
     CDS             complement(1. .1041)
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulator protein"
                     /protein_id="BAB69344.1"
                     /db_xref="GI:15824183"
                     /translation="MTEYDEQDEHHDLGRAARPGDFELPDVSGFDLLQYRESLTQGRR
                     PREETSQALLRLHLLHPVSEASDVCVPVSPSAAVDALVAPLEEERLKLRKKINDIRSA
                     MAAFDAVHSEVRSKEAVPLTLLVGGAAISASVTSSARASEREIITVQPGGSRPPELLA
                     ETMELNMSLLARGVRQRTLYQHSIRKHQLTMDYSRQVVAAGAEIRTVTEIFERLIICD
                     DTVAYIPTNDERANEALEIRHPALIRYLRAGFERAWEGAEPISPADPLPKRAREDAIL
                     VRIAQLLVDGHTDESISQRLGIGVRTVAKRISTLSGRLGSRSRGQLGYLIARSGMLDE
                     PSPPAPPAPPSG"
     CDS             1723. .2700
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB69345.1"
                     /db_xref="GI:15824184"
                     /translation="MSSWLPRNPAELDKAVTGAERDAILDAIRSLNDDARLGDSEAFY
                     AQQRLLARIYGLHTLIPEGPSAEGSVLLHTITRLLEEATIEAEDGYIEPGLIDQAPAE
                     PAAYLSWLKKLAREHRVYKHPYYREFIRDEATSADLRTYVMQESVVDGRFDDLLAMMQ
                     VGTSGAAKMEIAQNFWDEMGNGKPEEVHTHLFNKIFEVFEIPADELERSLTANALLSG
                     NLAVLLCRYRQYYPEAVGFLGMTEWLVPDRFVQVVHAWERLGLPDIGIVYHRLHITVD
```

```
                SQHAAGWPHNVVIPAAESEHMRRAIARGTLWRLNSSARYLDERMPAVAV"
     CDS        2732. .3427
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB69346.1"
                /db_xref="GI:15824185"
                /translation="MPLNEQKFALIELPEVSQELLDSYEDLPVDEYMGNGTRFKRFSQ
                YRLRPEGTEGWTFELLAHRDYTTFTKFNKVGGGIRRVYQPIKADFVPLISHGIEELGL
                DRSEDWQINVHQNRTRADGDYAGPLTPEGVHHDGHEFVMIAVLNKNNVAGAQTRLWEP
                GADAPFWTGTLEPGHAVLLDDRGLAHDVTDVESADGGPANRDIVIIAFSRWAEKWYGE
                EHDEAALADGAES"
     CDS        3474. .4397
                /codon_start=1
                /transl_table=11
                /product="putative diaminopimelate epimerase"
                /protein_id="BAB69347.1"
                /db_xref="GI:15824186"
                /translation="MGEHSTSSTVVTPASTPSASSPPAEFEESAEFFTYHALGNDYVV
                VDPRGSDFTPGPAAVRLLCDRHFGVGADGVLFGPLGAPEAGEPVRLRMFNSDGTECEK
                SGNGLRMFALHLAETYGEPWSGDEEFPLRTAAGDVPVRILDLASGTVQVGMGRPEFGA
                TGESLDVGGRRLAVTSVHNGNPHAVVPLAEISAGLARELGPLIAGHTRFPGRTNVQFL
                AIVDRRTIAIEVYERGAGYTLASGSSACAAAAAAAHALGLTGPDVEVRMPGGRIDVSIA
                ADGSVSMTGTAERISSGRFSPVFRNKLETAR"
     CDS        4432. .5535
                /codon_start=1
                /transl_table=11
                /product="putative aminotransferase"
                /protein_id="BAB69348.1"
                /db_xref="GI:15824187"
                /translation="MTTATTVSTYPMRRWVFDDAAGRYDIDLGDSNLPCGTLDELTPL
                GGLELDYGHNRGIPRLRALVADLYGTDDDHVLITHGAQQALYLVYNALLRPGDQVITF
                RPGWQQSWDMPAELGCRVDILDHRPDLTVDVEAAAAVAGPDLRLIVVNTPCNPTGQRV
                SDADLAALHTLAARTGAYLLLDEDYALELKDSAATRLERTISVSSLSKVYGFPGLRIG
                WLYGPPELVVACTERKFLSNIGNSVLCETLACDVLERRDHYTERYRTLTGEGLAVVRA
                WAARNTDAVRLTEPRGTPFAWFRLLTGEAPLDLCRRVLDTGVLLMPGETLGGPDGFRL
                CFARDPHLISEGLRRLDPVLRPQAVRNMTSGAV"
     CDS        5673. .8723
                /codon_start=1
                /transl_table=11
                /product="non-ribosomal peptide synthetase"
                /protein_id="BAB69349.1"
                /db_xref="GI:15824188"
                /translation="MAYRLRGTLDVGALHRALTSLVTRNTALRTTFRSAEGTPYGVVA
                PAGEVDLEIVDLADADLDATLRARSEAPFDLAAGPLYRFSLLRLSAEDHVLVLVFHHI
                NTDGWSSALINREISGAYNAFVQGTQPALPELEATYEDHVRAQREQHGGDSWEQQLKF
                WKGKLAGLPVVEMPADRARPATPTQNGETLTFDFPAELLHGARALAQQEGASLFMVLT
                TALNVVVGRYTGQEDIPLGIPMLGRTEPELEDVVGLFINMLVLRTDLSGDPTFAELIE
                RVADANMDVYDNQDVPFDQIVSHVQPAREAGRNPLFQIGTQLLGSGTGGSDLGLTGLT
                DEQLDVAVPRSRYDLSLTFLESADRLRFVAEYTTDLFDGIRIEAMARHIQNVLAAAVE
                DAEQPISKLAMLGEDERRLLLEAGRGPDFDFSHDPVHTTVLQVAERTPHAVAAVCRGV
                EMTYGELARRSGKLARHLRARGIRPGQVVGIVMDRDLDALVAMLGVMRAGGAYAVMDP
                SHPAARIDYMLQDTAAPVVITRADVAGGLPTSEDRKVVLIDADWAAIEAESDTEPLEE
                LADRETLAYVLYTSGSTGRPKGVMIEHRSLRVFIEAYRHTFGEFTPEDRLLQQPSLIF
                DMSQGEIYTALITGATLVLVSPEDASSPEALGTLIREQRVTYVGLSPTLLSLLDGGPY
                PHLKYIMGGAEVLPAELVNKWNIPGRKFVNLYGPTEAAIACTEYLCEHIEWRSAPPIG
                RPEYGRLHYVVDRHGNLAPVGVPGELLIGGEDGLARGYLNLPELTDEKFVPNPFEPTG
                RVYRTGDLVRWTPDYQVDFIGRIDNQVKLRGLRIELGEIESALVAHPQIRMALVLLKA
                DPQGEKQLVGYYTVLGDQAPTLAELRRHLGESMPEYMVPTAWVKLDEFPLTLAKKIDR
                HALPEPSYGADGTEGRAYVAPATPTEERVAEAFATVLKADQISADASFFELGGNSLQA
                MRAVSRINKAFGVKINVRMLYGGATVGGVAEAIDELLAAKAGRD"
     CDS        8737. .9549
                /codon_start=1
                /transl_table=11
                /product="non-ribosomal peptide synthetase"
                /protein_id="BAB69350.1"
                /db_xref="GI:15824189"
                /translation="MPLREDGTHPPLYCVHPVSGSGYVYAGLAELLGPQQPSYGFEAP
                GFDDGGAPLNSIEALAERHLASLRSARPRGPYLLLGWSLGGVVAYHMAQRLVEAGEEV
                PLLVLIDATVPQPFPLPPEKALLHAFLRDLVTAMGEPPEALDALLTPYPDDVTPEAVW
                TAVTAAGDEVLPEEFDEEFLQERYDVFRAHAEALYAYEHRGPYAGRVLFLKAADSLDA
                RTRWETLAPRVERHTITGDHDSIWAGEAFATLAGLVRQGLAQARSALAPAEG"
     CDS        9609. .10901
                /codon_start=1
                /transl_table=11
                /product="putative export protein"
                /protein_id="BAB69351.1"
                /db_xref="GI:15824190"
                /translation="MEDTAAAGSAAPRRRGGGLWRNRDFLVFWTGETLSLYGTQVTTL
                ALPLAAVLVFDASAEELGVLRFLQLVPFLGLAMLFGVWVDRVRKRPVMIGANAARMVL
                IGLVPALSAADVLNLPLLYAVAFGIGVAQVLFDVSWMSYVPVLVKDRRHLVEANTKLS
                TTSSSADVAGPGVAGVLVNALTAPTAMAVNAVTYLLSIVSLLAIRTPEPRPAKPERRR
                SIRADLVEGLRFVFGSRLLRAIALVGCACNFFSVMILSLFLLYAVRGQHLSAATIGLV
                LGVGAVGGLAGSALSGTVIRRVSVGKAYLVSVSGIFVAPLLIPAAGGSKLMVTGVFIA
                SFLVFYASLSVANVVIVSLRQTITPDPLMARMTAAMRTLMHGGGALGGPVGGLLAGAL
                GLHEALWLAAVLSAVMLVPIALSPVSRLHEMPAPAPVP"
     CDS        10979. .12856
                /codon_start=1
                /transl_table=11
                /product="non-ribosomal peptide synthetase"
                /protein_id="BAB69352.1"
                /db_xref="GI:15824191"
                /translation="MSPDVDEIAAWNGVSPHVDEIAAWNDTARELPGGGDTLPELFRT
                QVSRTPDATALVFRDTRLSYTELDERSDRLARLLAGHGAAPDRVVALTVPRSVELVVA
                LLAVLKTGAAYLPVDPDYPADRIGYLLADAAPVLLLTHSDVEATLPPGARPEGLPRLV
                VDAQEEPLERAPFAPARPPAAQDAAYVIYTSGSTGRPKGVVVPHAGIVNRLHWMQHEY
                RLTGDDRVLQKTPSGFDVSVWEFFWPLITGATLVVAEPGGHKDPAYLAELIRGERITT
                VHFVPSMLQVFLDEPAAALCTGLRRVVCSGEALPVELADRFTASLPGVPLHNLYGPTE
                ASVDVTYWEYTPEPGATSVPIGRPVWNTRLYVLDAALRPAAVGETGELYLAGVQLALG
                YLNRPGLTAERFVADPYGPPGSRMYRTGDLARRRPDGALDYVGRADHQVKIRGLRVEL
                GEIEAALARHPAVSGAAVLVREDRPGDKRLVGYVVPAPGGTVDPAELRTLAATTLPDY
                MVPSTVVVLESLPLTPNGKLDRKALPAPPDAAAAPARVAPRTPQEALLCDLFAEVLGL
                SETQVGVEDSFLELGGHSLSAIRLVSRIRAAFDTDLKVRQVFKTPTVAALAACLASAG
                R"
     CDS        complement(13245. .14198)
                /codon_start=1
                /transl_table=11
                /product="ABC transporter ATP-binding protein BldKE"
                /protein_id="BAB69353.1"
                /db_xref="GI:15824192"
                /translation="MQAVDGLDFQVAEGESLGLVGESGCGKSTTGRLITRLLEPTSGK
                ISYRGQDITHADRKKLAPVRSEIQMIFQDPYASLNPRHTVGKIIAGPMEINGIQPSGG
                REKRVRELLEIVGLNPEHYNRFPHEFSGGQRQRIGVARALALEPKLIVADEPVSALDV
                SIQAQVVNLLQDLQKDLGIAFVFIAHDLAIVRHFSQRVAVMYLGKIVEIADRADLYEN
                PRHPYTRALLSAVPEATVEDGEARERIRLTGDVPSPINPPSGCRFRTRCWKATDKCAS
                EAPPLVQVEGNNAGHLTACHYPETKETVPAPRLSKDPKAAV"
     CDS        complement(14315. .15385)
                /codon_start=1
                /transl_table=11
                /product="ABC transporter intracellular ATPase BldKD"
                /protein_id="BAB69354.1"
                /db_xref="GI:15824193"
                /translation="MKSTATKPAGERNPQATGPPFLSVQDLKVHFSTEDGTVKAVDGLS
                FDLERGKTLGIVGESGSGKSVTNLTILGLHNPDSTAVEGSITLDGEELNGASEKTLEK
                LRGNKMAMIFQDALTSLSPYHTIGKQIGETYRKHTGCSKAAARARAIEMLTRVGIPQP
                DVRVDDYPHQFSGGMRQRAMIAMALVCDPELLIADEPTTALDVTVQAQIMDLLKDLQK
                EFGTAIIFITHDLGVIADIADDVLVMYGGRCVERGTKKEVLRTPQHPYTWGLLGSMPS
                LEGPVDVPLTPIPGSPPSLLNPPSGCRFHPRCSFAEKVEDGRCSSERPPLEITDGRGA
                ACHLTADQRRAFFADLAASPAN"
     CDS        complement(15446. .16339)
                /codon_start=1
                /transl_table=11
                /product="transport membrane protein BldKC"
                /protein_id="BAB69355.1"
                /db_xref="GI:15824194"
                /translation="MPRDPARLACGKLCTPETLEVIRHNMGITDSLPVQYWHWLTGIF
                LGRDYGSYGHCDAPCLGYSFVNRESVLGTIMDRFPTTASLALGSSVVFLIGGIGAGMI
                AAVKQGKALDKVASGASLVASSMQIYFVGTLALYCFVYQWQILDQPSYTDFTKNPASW
                ANGLLLPWLVLSLIWTANYTRMTRSQLVEQLNEDYVRTARAKGLSRRNVFFRFAWRGA
                MGPIVTIFGIDLGVLLGGAMITEKTFSLQGIGMLAIKAVTNNDLPMLLGIMVITASFV
                VIANIVVDAVYALIDPRIRLA"
     CDS        complement(16552. .18315)
                /codon_start=1
                /transl_table=11
                /product="ABC transport protein BldKB"
```

```
  Query Match            12.1%;  Score 121.4;  DB 1;  Length 23451;
  Best Local Similarity  47.2%;  Pred. No. 0.036;
  Matches  434;  Conservative    0;  Mismatches  481;  Indels    4;  Gaps    2;

Qy         54 GACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCG 113
              | |||   |  ||  | |  || |        ||||| |  || ||  |  ||||    |
Db       8286 GGCGACCAGGCCCCGACCCTCGCCGAACTCCGCCGCCACCTCGGCGAGTCGATGCCCGAG 8345

Qy        114 GTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTG 173
              |  || |  ||  | |     || |  |      | ||| ||| |   |||    |
Db       8346 TACATGGTGCCGACCGCCTGGGTGAAGCTGGACGAGTTCCCGCTCACGCTCGCCAAGAAG 8405

Qy        174 AACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCG 233
              | | | |      || |      |           ||  | ||||  | |   | |    |
Db       8406 ATCGACCGGCACGCGCTGCCCGAGCCGTCGTACGGCGCCGACGGTACGGAAGGCCGGGCC 8465

Qy        234 TCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCC-CCGGCGGCCTCCGC 292
              | || ||| |||  | |  | |||  |  |  | |||||  ||||  || |     |||
Db       8466 TACGTCGCCCCCGCCACGCCCACCGAGGAGCGGGTCGCCGAGGCCTTCGCCACGGTCCTG 8525

Qy        293 AGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCC 352
              |   |||        || |||    || | || ||  |   | | |||||  |      ||
Db       8526 AAGGCCGATCAGATCAGCGCGGACGCCAGCTTCTTCGAACTCGGCGGCAACTCGCTCCAG 8585

Qy        353 GGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGA 412
              |     |     | |  |||   |  |   | | || ||          |  ||  |
Db       8586 GCGATGCGGGCGGTCAGCCGGATCAACAAGGCCTTCGGCGTGAAGATCAACGTCCGGATG 8645

Qy        413 GGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCC 472
              || ||||| |     ||| | |||| |     |  ||||    | || || ||    | |||
Db       8646 CTCTACGGCGGCGCCACCGTGGGCGGCGTGGCCGAGGCCATCGACGAACTGCTCGCCGCC 8705

Qy        473 GCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACG 532
              |||   | | | |  || |   |   || ||      ||  |  |  |||| | |
Db       8706 AAGGCC---GGCCGTGACTGACTCGACCACCCTGGTGCCGCTCCGCGAGGACGGCACCCA 8762

Qy        533 CGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTC 592
              |  |      || ||| ||||||    |    | ||     |    |   |  |       ||  |
Db       8763 CCCGCCGCTGTACTGCGTGCACCCGGTCTCCGGTTCGGGCTACGTGTACGCGGGGCTCGC 8822

Qy        593 TGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCT 652
              |      |||||| |  |  | ||||  | || |||| |||| | |||  ||||||
Db       8823 CGAACTCCTCGGCCCTCAGCAGCCGTCGTACGGCTTCGAGGCGCCGGGGTTCGACGACGG 8882

Qy        653 CGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCA 712
              ||    |||| |||     |||||||  | ||||||  |   || | || | |||||||
Db       8883 CGGGGCGCCCCTCAACTCGATCGAGGCGCTCGCGGAACGGCACCTGGCGTCCCTGCGGTC 8942

Qy        713 GCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTT 772
              |    | ||| | || || ||| |  ||    ||   |||   ||||||  |  ||||||
Db       8943 GGCCCGGCCGCGGGGACCGTACCTGCTGCTCGGCTGGTCGCTCGGCGGTGTGGTCGCCTA 9002

Qy        773 CGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGA 832
              | | ||||||| || |||| |   |   || |||   ||  | | ||| ||| |  ||||
Db       9003 CCACATGGCCCAGCGGCTCGTGGAGGCCGGCGAGGAGGTGCCGCTGCTCGTCCTCATCGA 9062

Qy        833 CACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCT 892
              | |       |     ||  ||    || |  |  |  ||       |  || |  ||||
Db       9063 CGCGACGGTGCCCCAGCCGTTCCCGCTGCCGCCGGAGAAGGCGCTGCTGCACGCCTTCCT 9122

Qy        893 GCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGAT 952
              || |  |||  | | ||| |                  |  |  ||||       | ||
Db       9123 GCGGGACCTCGTGACGGCGATGGGGGAGCCGCCGGAGGCGCTGGACGCGCTGCTGACGCC 9182

Qy        953 CAAGCCCGACGACCTCGCC 971
              | | |||||||||| || ||
Db       9183 CTACCCCGACGACGTCACC 9201


RESULT 3
BA000030_39/c
WPCOMMENT
  Sequence split into 91 fragments  LOCUS BA000030 Accession BA000030
```

Fragment Name	Begin	End
BA000030_00	1	110000
BA000030_01	100001	210000
BA000030_02	200001	310000
BA000030_03	300001	410000
BA000030_04	400001	510000
BA000030_05	500001	610000
BA000030_06	600001	710000
BA000030_07	700001	810000
BA000030_08	800001	910000
BA000030_09	900001	1010000
BA000030_10	1000001	1110000
BA000030_11	1100001	1210000
BA000030_12	1200001	1310000
BA000030_13	1300001	1410000
BA000030_14	1400001	1510000
BA000030_15	1500001	1610000
BA000030_16	1600001	1710000
BA000030_17	1700001	1810000
BA000030_18	1800001	1910000
BA000030_19	1900001	2010000
BA000030_20	2000001	2110000
BA000030_21	2100001	2210000
BA000030_22	2200001	2310000
BA000030_23	2300001	2410000
BA000030_24	2400001	2510000
BA000030_25	2500001	2610000
BA000030_26	2600001	2710000
BA000030_27	2700001	2810000
BA000030_28	2800001	2910000
BA000030_29	2900001	3010000
BA000030_30	3000001	3110000
BA000030_31	3100001	3210000
BA000030_32	3200001	3310000
BA000030_33	3300001	3410000
BA000030_34	3400001	3510000
BA000030_35	3500001	3610000
BA000030_36	3600001	3710000
BA000030_37	3700001	3810000
BA000030_38	3800001	3910000
BA000030_39	3900001	4010000
BA000030_40	4000001	4110000
BA000030_41	4100001	4210000
BA000030_42	4200001	4310000
BA000030_43	4300001	4410000
BA000030_44	4400001	4510000
BA000030_45	4500001	4610000
BA000030_46	4600001	4710000
BA000030_47	4700001	4810000
BA000030_48	4800001	4910000
BA000030_49	4900001	5010000
BA000030_50	5000001	5110000
BA000030_51	5100001	5210000
BA000030_52	5200001	5310000
BA000030_53	5300001	5410000
BA000030_54	5400001	5510000
BA000030_55	5500001	5610000
BA000030_56	5600001	5710000
BA000030_57	5700001	5810000
BA000030_58	5800001	5910000
BA000030_59	5900001	6010000
BA000030_60	6000001	6110000
BA000030_61	6100001	6210000
BA000030_62	6200001	6310000
BA000030_63	6300001	6410000
BA000030_64	6400001	6510000
BA000030_65	6500001	6610000
BA000030_66	6600001	6710000
BA000030_67	6700001	6810000
BA000030_68	6800001	6910000
BA000030_69	6900001	7010000
BA000030_70	7000001	7110000
BA000030_71	7100001	7210000

```
BA000030_72       7200001   7310000
BA000030_73       7300001   7410000
BA000030_74       7400001   7510000
BA000030_75       7500001   7610000
BA000030_76       7600001   7710000
BA000030_77       7700001   7810000
BA000030_78       7800001   7910000
BA000030_79       7900001   8010000
BA000030_80       8000001   8110000
BA000030_81       8100001   8210000
BA000030_82       8200001   8310000
BA000030_83       8300001   8410000
BA000030_84       8400001   8510000
BA000030_85       8500001   8610000
BA000030_86       8600001   8710000
BA000030_87       8700001   8810000
BA000030_88       8800001   8910000
BA000030_89       8900001   9010000
BA000030_90       9000001   9025608
Continuation (40 of 91) of BA000030 from base 3900001 (BA000030 Streptomyces avermitilis

  Query Match            12.1%;  Score 121.4;  DB 1;  Length 110000;
  Best Local Similarity  47.2%;  Pred. No. 0.019;
  Matches  434;  Conservative    0;  Mismatches  481;  Indels    4;  Gaps    2;

Qy         54 GACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCG 113
Db      34876 GGCGACCAGGCCCCGACCCTCGCCGAACTCCGCCGCCACCTCGGCGAGTCGATGCCCGAG 34817

Qy        114 GTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTG 173
Db      34816 TACATGGTGCCGACCGCCTGGGTGAAGCTGGACGAGTTCCCGCTCACGCTCGCCAAGAAG 34757

Qy        174 AACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCG 233
Db      34756 ATCGACCGGCACGCGCTGCCCGAGCCGTCGTACGGCGCCGACGGTACGGAAGGCCGGGCC 34697

Qy        234 TCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCC-CCGGCGGCCTCCGC 292
Db      34696 TACGTCGCCCCCGCCACGCCCACCGAGGAGCGGGTCGCCGAGGCCTTCGCCACGGTCCTG 34637

Qy        293 AGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCC 352
Db      34636 AAGGCCGATCAGATCAGCGCGGACGCCAGCTTCTTCGAACTCGGCGGCAACTCGCTCCAG 34577

Qy        353 GGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGA 412
Db      34576 GCGATGCGGGCGGTCAGCCGGATCAACAAGGCCTTCGGCGTGAAGATCAACGTCCGGATG 34517

Qy        413 GGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCC 472
Db      34516 CTCTACGGCGGCGCCACCGTGGGCGGCGTGGCCGAGGCCATCGACGAACTGCTCGCCGCC 34457

Qy        473 GCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACG 532
Db      34456 AAGGCC---GGCCGTGACTGACTCGACCACCCTGGTGCCGCTCCGCGAGGACGGCACCCA 34400

Qy        533 CGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTC 592
Db      34399 CCCGCCGCTGTACTGCGTGCACCCGGTCTCCGGTTCGGGCTACGTGTACGCGGGGCTCGC 34340

Qy        593 TGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCT 652
Db      34339 CGAACTCCTCGGCCCTCAGCAGCCGTCGTACGGCTTCGAGGCGCCGGGGTTCGACGACGG 34280

Qy        653 CGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCA 712
Db      34279 CGGGGCGCCCCTCAACTCGATCGAGGCGCTCGCGGAACGGCACCTGGCGTCCCTGCGGTC 34220

Qy        713 GCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTT 772
Db      34219 GGCCCGGCCGCGGGGACCGTACCTGCTGCTCGGCTGGTCGCTCGGCGGTGTGGTCGCCTA 34160

Qy        773 CGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGA 832
Db      34159 CCACATGGCCCAGCGGCTCGTGGAGGCCGGCGAGGAGGTGCCGCTGCTCGTCCTCATCGA 34100

Qy        833 CACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCT 892
Db      34099 CGCGACGGTGCCCCAGCCGTTCCCGCTGCCGCCGGAGAAGGCGCTGCTGCACGCCTTCCT 34040

Qy        893 GCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGAT 952
Db      34039 GCGGGACCTCGTGACGGCGATGGGGGAGCCGCCGGAGGCGCTGGACGCGCTGCTGACGCC 33980

Qy        953 CAAGCCCGACGACCTCGCC 971
Db      33979 CTACCCCGACGACGTCACC 33961


RESULT 4
AR621890
LOCUS       AR621890               4296 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 3291 from patent US 6833447.
ACCESSION   AR621890
VERSION     AR621890.1  GI:59726380
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4296)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 3291 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .4296
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            12.0%;  Score 120.2;  DB 6;  Length 4296;
  Best Local Similarity  52.2%;  Pred. No. 0.09;
  Matches  314;  Conservative    0;  Mismatches  283;  Indels    4;  Gaps    2;

Qy        326 CGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGC 385
Db       3605 CGCTGGGCGCGGACCTGACGCTGGCCACGCTGTTCGAGGCTCCCACGCTGGAGCAGTGCG 3664

Qy        386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTC-CGGCGCCTGATCCGCGCGCGGGGGACG 444
Db       3665 CCGCGCTGGTGCGTGAGGCCGCCGGGATTCCCTTCACGCCGGACGCGGGTCCGGGTGACG 3724

Qy        445 CACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGC 504
Db       3725 CGGCGCCGGCTGCGCAGGCCCCGGGCGTCAAGGCGGTGCCGAAG---AGCTGGTCTCCGC 3781

Qy        505 TCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTG 564
Db       3782 TGGTCACCATCCAGAAGGGCGGCAATACCCTCCCCTTCTTCTGCGTGCACGGCGCGGGCG 3841

Qy        565 GGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCG 624
Db       3842 GCAACGTCCTCAACTTCCGCGAGCTGGCGCAGACGCTCGGCAAGGAGCAGCCCTTCTACG 3901

Qy        625 CCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCG 684
Db       3902 GGCTCCAGGCGCGTGGCGTGGACGGCAAGCTGCCTCCGGCCGACAGCATCGAGGAGATGG 3961

Qy        685 CGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTG 744
Db       3962 CGTCCATCTACCTGGAGAGCATCCGTCAGGTGCAGCCCCACGGGCCCTACCTGCTCGGCG 4021

Qy        745 GGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGG 804
Db       4022 GCTACTCCGGCGGCGGCGTGGTCGCCTACAAAATAGCGCAGCGGCTGCACGCGCTGGGCG 4081
```

```
Qy        805 AGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGG 864
              ||| ||  | | | | |   | || |||||||  | | ||  | |           |
Db       4082 AGCCGGTGGCGCTGGTCGCGTTCCTGGACACCTTCCACCCGTCCACCCAGGAGCGACGCC 4141

Qy        865 AGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCA 924
              || | |      ||   |   | |   ||||  |     |  | || || || | ||| |
Db       4142 AGTCCCTCCATGAGCGCCTGCGGGAGCTGCGCCGGGAGGGCGCGGGCAGCTACGTCAGCC 4201

Qy        925 G 925
              |
Db       4202 G 4202


RESULT 5
AR619757/c
LOCUS       AR619757               17727 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 1152 from patent US 6833447.
ACCESSION   AR619757
VERSION     AR619757.1  GI:59724163
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17727)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE     Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 1152 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .17727
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             12.0%;  Score 120.2;  DB 6;  Length 17727;
  Best Local Similarity   52.2%;  Pred. No. 0.05;
  Matches  314;  Conservative    0;  Mismatches  283;  Indels    4;  Gaps    2;

Qy        326 CGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGC 385
              ||| |  | ||   |   ||| | ||  || |     |   ||| ||| |   | | |
Db      14123 CGCTGGGCGCGGACCTGACGCTGGCCACGCTGTTCGAGGCTCCCACGCTGGAGCAGTGCG 14064

Qy        386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTC-CGGCGCCTGATCCGCGCGCGGGGGACG 444
              |  ||| |   ||   |  | ||  ||  | |    |||| ||  || |  |||| ||||
Db      14063 CCGCGCTGGTGCGTGAGGCCGCCGGGATTCCCTTCACGCCGGACGCGGGTCCGGGTGACG 14004

Qy        445 CACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGC 504
              |   ||| |  ||| |   |   | |    || | | ||||  |   | |          ||
Db      14003 CGGCGCCGGCTGCGCAGGCCCCGGGCGTCAAGGCGGTGCCGAAG---AGCTGGTCTCCGC 13947

Qy        505 TCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTG 564
              | |||  |||| |   | |||||      |      ||  ||||| ||||||||||||   |
Db      13946 TGGTCACCATCCAGAAGGGCGGCAATACCCTCCCCTTCTTCTGCGTGCACGGCGCGGGCG 13887

Qy        565 GGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCG 624
              |   ||  |      ||      ||| |     ||||||||| |||  | || |||| ||
Db      13886 GCAACGTCCTCAACTTCCGCGAGCTGGCGCAGACGCTCGGCAAGGAGCAGCCCTTCTACG 13827

Qy        625 CCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCG 684
                |||||||  || |||||  |||||          ||   |   |||||||||||||| |
Db      13826 GGCTCCAGGCGCGTGGCGTGGACGGCAAGCTGCCTCCGGCCGACAGCATCGAGGAGATGG 13767

Qy        685 CGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTG 744
              || ||   ||| | |   || | || |||      ||     ||||| ||| |  | || |
Db      13766 CGTCCATCTACCTGGAGAGCATCCGTCAGGTGCAGCCCCACGGGCCCTACCTGCTCGGCG 13707

Qy        745 GGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGG 804
              |   |||  |||||||  | ||||||  | | || || | ||  ||| | | |||  || |
Db      13706 GCTACTCCGGCGGCGGCGTGGTCGCCTACAAAATAGCGCAGCGGCTGCACGCGCTGGGCG 13647

Qy        805 AGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGG 864
              |||  ||  | | | | |   | || |||||||  | | ||  | |           |
Db      13646 AGCCGGTGGCGCTGGTCGCGTTCCTGGACACCTTCCACCCGTCCACCCAGGAGCGACGCC 13587

Qy        865 AGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCA 924
              || | |      ||   |   | |   ||||  |     |  | || || || | ||| |
Db      13586 AGTCCCTCCATGAGCGCCTGCGGGAGCTGCGCCGGGAGGGCGCGGGCAGCTACGTCAGCC 13527

Qy        925 G 925
              |
Db      13526 G 13526


RESULT 6
BA000030_72/c
WPCOMMENT
  Sequence split into 91 fragments  LOCUS BA000030 Accession BA000030
     Fragment Name      Begin      End
     BA000030_00            1    110000
     BA000030_01       100001    210000
     BA000030_02       200001    310000
     BA000030_03       300001    410000
     BA000030_04       400001    510000
     BA000030_05       500001    610000
     BA000030_06       600001    710000
     BA000030_07       700001    810000
     BA000030_08       800001    910000
     BA000030_09       900001   1010000
     BA000030_10      1000001   1110000
     BA000030_11      1100001   1210000
     BA000030_12      1200001   1310000
     BA000030_13      1300001   1410000
     BA000030_14      1400001   1510000
     BA000030_15      1500001   1610000
     BA000030_16      1600001   1710000
     BA000030_17      1700001   1810000
     BA000030_18      1800001   1910000
     BA000030_19      1900001   2010000
     BA000030_20      2000001   2110000
     BA000030_21      2100001   2210000
     BA000030_22      2200001   2310000
     BA000030_23      2300001   2410000
     BA000030_24      2400001   2510000
     BA000030_25      2500001   2610000
     BA000030_26      2600001   2710000
     BA000030_27      2700001   2810000
     BA000030_28      2800001   2910000
     BA000030_29      2900001   3010000
     BA000030_30      3000001   3110000
     BA000030_31      3100001   3210000
     BA000030_32      3200001   3310000
     BA000030_33      3300001   3410000
     BA000030_34      3400001   3510000
     BA000030_35      3500001   3610000
     BA000030_36      3600001   3710000
     BA000030_37      3700001   3810000
     BA000030_38      3800001   3910000
     BA000030_39      3900001   4010000
     BA000030_40      4000001   4110000
     BA000030_41      4100001   4210000
     BA000030_42      4200001   4310000
     BA000030_43      4300001   4410000
     BA000030_44      4400001   4510000
     BA000030_45      4500001   4610000
     BA000030_46      4600001   4710000
     BA000030_47      4700001   4810000
     BA000030_48      4800001   4910000
     BA000030_49      4900001   5010000
     BA000030_50      5000001   5110000
     BA000030_51      5100001   5210000
     BA000030_52      5200001   5310000
     BA000030_53      5300001   5410000
     BA000030_54      5400001   5510000
     BA000030_55      5500001   5610000
     BA000030_56      5600001   5710000
```

BA000030_57	5700001	5810000
BA000030_58	5800001	5910000
BA000030_59	5900001	6010000
BA000030_60	6000001	6110000
BA000030_61	6100001	6210000
BA000030_62	6200001	6310000
BA000030_63	6300001	6410000
BA000030_64	6400001	6510000
BA000030_65	6500001	6610000
BA000030_66	6600001	6710000
BA000030_67	6700001	6810000
BA000030_68	6800001	6910000
BA000030_69	6900001	7010000
BA000030_70	7000001	7110000
BA000030_71	7100001	7210000
BA000030_72	7200001	7310000
BA000030_73	7300001	7410000
BA000030_74	7400001	7510000
BA000030_75	7500001	7610000
BA000030_76	7600001	7710000
BA000030_77	7700001	7810000
BA000030_78	7800001	7910000
BA000030_79	7900001	8010000
BA000030_80	8000001	8110000
BA000030_81	8100001	8210000
BA000030_82	8200001	8310000
BA000030_83	8300001	8410000
BA000030_84	8400001	8510000
BA000030_85	8500001	8610000
BA000030_86	8600001	8710000
BA000030_87	8700001	8810000
BA000030_88	8800001	8910000
BA000030_89	8900001	9010000
BA000030_90	9000001	9025608

Continuation (73 of 91) of BA000030 from base 7200001 (BA000030 Streptomyces avermitilis

Query Match 11.2%; Score 112.4; DB 1; Length 110000;
Best Local Similarity 49.0%; Pred. No. 0.1;
Matches 361; Conservative 0; Mismatches 366; Indels 9; Gaps 2;

```
Qy      219 GCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCC 278
            || || |      ||  |||||||  ||    |||| || |  |  |  ||    ||| ||||
Db    51802 GCCGCGCCCGCGCCCGCCGCCGCCCCGGCCGCCCCGGCTCCCGCCCCGGCTGCCCCGGCT 51743

Qy      279 CCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCG 338
            |||||  ||  | ||  |  |  |||||  || ||||||||||     ||||  || | |||
Db    51742 CCGGCCGCTCCGGCTCCGGCTCCGGCCGCCCCGGCCCCGGCTCAGGCCGCCCCCGCGGCG 51683

Qy      339 GCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCG 398
             || ||||  |  |   |||| |  |||||||  || ||  |   |||| | || | ||||
Db    51682 CCCGCCGCCCCGGCTCAGGCCGCTCCGGCCCCGGCCGCTCCGCAGGCCCCGGCCGCCCCG 51623

Qy      399 GCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCG 458
            |||  ||  || ||   | |||||  | |  ||   ||||     ||       ||| ||
Db    51622 GCGCCGGCCCCCGTCACCCCGGCCGCGGCGCCCGCCGCGACCTCCGGTGACGACGGCGCC 51563

Qy      459 GACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAAC 518
             ||   | | | |  || | |||  |   || |  || ||    | | | || |  |
Db    51562 TACGTGACCCCGCTGGTGCGCAAGCTCGCCGCCGAGAACGGCGTCGACCTGGCCACCGTC 51503

Qy      519 CCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCG 578
              | ||  |     |||  ||    || |   ||| | ||   | |    |||  || | |
Db    51502 AAGGGCACCGGCGTCGGCGGCCGCATCCGCAAGCAGGACGTCACCGCCGCCGCCGAGGCC 51443

Qy      579 CTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGG 638
                |  ||     || | |||| || | | | |   |||  |  |||     |||||| |
Db    51442 GCGAAGGCCGCCGCTCCGGCGCCCGCCGCCGCT---CCGGCCGCCGCGAAGAAGGCCCCG 51386

Qy      639 GGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTC 698
            | | | || |    | || | | |  |  || | || ||  |  |      ||   | |||
Db    51385 GCCCTGGAGGTCTCCCCCCTCCGCGGCCAGACCGTCAAGATGCCCCGCATCCGCAAGGTC 51326

Qy      699 GCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGC 758
```

```
                | ||       |    |    ||  |  |   |||    ||    ||| ||||| | |
Db    51325 ATCGGC------GACAACATGGTGAAGGCCCTGCACGAGCAGGCCCAGCTCTCCTCGGTC 51272

Qy      759 GGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGG 818
            | |    | | ||          | |||||| | |||| | |  ||  |  || |  |
Db    51271 GTCGAGGTGGACATCACCCGCCTGATGAAGCTCCGCGCGCAGGCGAAGGACTCCTTCGCG 51212

Qy      819 CTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGAC 878
                |  |  || ||       || || |  ||  | ||| | ||| | || |    |
Db    51211 GCCCGCGAGGGCGTCAAGCTCTCCCCGATGCCGTTCTTCGTGAAGGCGGCGGCCCAGGCG 51152

Qy      879 ATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGAC 938
              |  |  |  || | |  ||     |  |  ||||     |    ||         | ||
Db    51151 CTGAAGGCCCACCCGGCCGTCAACGCCCGGATCAACGTGGACGAGGGCACGATCACCTAC 51092

Qy      939 GGAGACACGGCGATCA 954
              |||||  | || ||
Db    51091 TTCGACACCGAGAACA 51076
```

RESULT 7
CP000010_11
WPCOMMENT
Sequence split into 36 fragments LOCUS CP000010 Accession CP000010

Fragment Name	Begin	End
CP000010_00	1	110000
CP000010_01	100001	210000
CP000010_02	200001	310000
CP000010_03	300001	410000
CP000010_04	400001	510000
CP000010_05	500001	610000
CP000010_06	600001	710000
CP000010_07	700001	810000
CP000010_08	800001	910000
CP000010_09	900001	1010000
CP000010_10	1000001	1110000
CP000010_11	1100001	1210000
CP000010_12	1200001	1310000
CP000010_13	1300001	1410000
CP000010_14	1400001	1510000
CP000010_15	1500001	1610000
CP000010_16	1600001	1710000
CP000010_17	1700001	1810000
CP000010_18	1800001	1910000
CP000010_19	1900001	2010000
CP000010_20	2000001	2110000
CP000010_21	2100001	2210000
CP000010_22	2200001	2310000
CP000010_23	2300001	2410000
CP000010_24	2400001	2510000
CP000010_25	2500001	2610000
CP000010_26	2600001	2710000
CP000010_27	2700001	2810000
CP000010_28	2800001	2910000
CP000010_29	2900001	3010000
CP000010_30	3000001	3110000
CP000010_31	3100001	3210000
CP000010_32	3200001	3310000
CP000010_33	3300001	3410000
CP000010_34	3400001	3510000
CP000010_35	3500001	3510148

Continuation (12 of 36) of CP000010 from base 1100001 (CP000010 Burkholderia mallei ATCC

Query Match 11.0%; Score 110; DB 1; Length 110000;
Best Local Similarity 48.8%; Pred. No. 0.16;
Matches 296; Conservative 0; Mismatches 310; Indels 0; Gaps 0;

```
Qy      308 CGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGC 367
            ||  || | ||   |   ||| | |   || | |   |    ||     || ||  ||  |
Db    65396 CGACGCGCTGGGCGTGACCGTACCGGTCGAGTCGATCCTCGCGAGTCCGACCATCGTCGA 65455

Qy      368 CCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGA 427
              |  || |  | ||  |          | ||  ||   ||| ||| |   | |||||  |
```

```
Db      65456 GCTTGCCGGATACGTCGAGCAATTGCGCTCGCGGGACGTCCGCGACGGCGCTGCGCCCGT 65515

Qy        428 TCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGG 487

Db      65516 GTCGATCGGCGAAACACCGGCGGACGCGGATGCTCGAGCTCAGGCGCAGGCGGATACGGT 65575

Qy        488 GCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTG 547

Db      65576 GCACACCGATTGCCTGATCGTCATTCAAGCAGGCGGCGCCGAACAAGCGCCGGTGTTCTG 65635

Qy        548 GGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCAC 607

Db      65636 CATCCCGGGCGCGGGGGGCAGCGTCGCGTCGTTCGTTGCGCTTGCGAGCATGCTGCGCGC 65695

Qy        608 GGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCAC 667

Db      65696 CGACATACCGGTATACGGCTTGCAGCCTCGCGGGCTGGACGGCCTGGGGCCGCCGGACCG 65755

Qy        668 GAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGG 727

Db      65756 ATCCGTCGAAGCGGCTGCGCGCCGGTACGCGCGGGCCATTCTGGATGCCACCCCGCCCGG 65815

Qy        728 GCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCA 787

Db      65816 GCCGCCGCGCATCGTCGGCCACTCGTTCGGCGGCTGGATCGCGCTCGAGACAGCGCGGCT 65875

Qy        788 GCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGT 847

Db      65876 GCTGGACGGCATGGGAACGCGCTGCGCCCCGCTCGTCCTGCTCGATTCGAATCCACCGCC 65935

Qy        848 CGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCAT 907

Db      65936 CGCGTCGCAGGCCTGGCGCGCGCCTTCCGAGGCAGACATGCTGCGCACGCTCGTCGGCCT 65995

Qy        908 GGCCAA 913

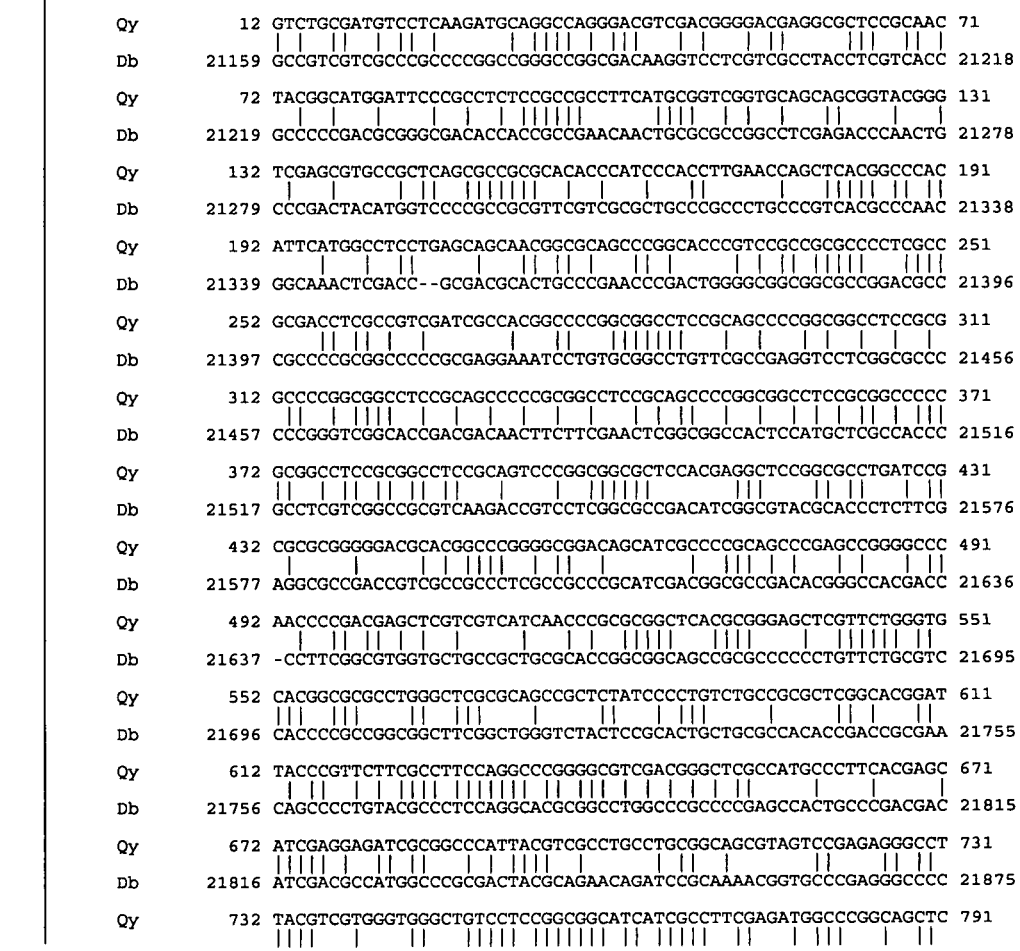
Db      65996 GCTCGA 66001
```

```
RESULT 8
SPSNBCDE
LOCUS       SPSNBCDE               22449 bp    DNA     linear   BCT 18-APR-2005
DEFINITION  S.pristinaespiralis snbC gene & snbDE gene.
ACCESSION   Y11548
VERSION     Y11548.1  GI:2052247
KEYWORDS    Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbC
            gene; snbDE gene.
SOURCE      Streptomyces pristinaespiralis
  ORGANISM  Streptomyces pristinaespiralis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   de Crecy-Lagard,V.A., Saurin,W., Thibaut,D., Gil,P., Naudin,L.,
            Crouzet,J. and Blanc,V.
  TITLE     Streptogramin B biosynthesis in Streptomyces pristinaespiralis and
            Streptomyces virginiae: molecular characterization of the last
            structural peptide synthase gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 22449)
  AUTHORS   de Crecy-Lagard,V.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-1997) V.A. de Crecy-Lagard, Institut Pasteur,
            Unit de Biochimie Microbienne, 28 rue du Dr Roux, 75724 Paris
            cedex, FRANCE
FEATURES             Location/Qualifiers
     source          1. .22449
                     /organism="Streptomyces pristinaespiralis"
                     /mol_type="genomic DNA"
                     /strain="SP92"
                     /db_xref="taxon:38300"
                     /clone_lib="pIBV3 cosmid"
     gene            115. .7890
                     /gene="snbC"
     CDS             115. .7890
                     /gene="snbC"
                     /codon_start=1
                     /transl_table=11
                     /product="Pristinamycin I synthase 2"
                     /protein_id="CAA72311.1"
                     /db_xref="GI:2052248"
                     /db_xref="GOA:Q54959"
                     /db_xref="InterPro:IPR000873"
                     /db_xref="InterPro:IPR001242"
                     /db_xref="InterPro:IPR006163"
                     /db_xref="InterPro:IPR010060"
                     /db_xref="InterPro:IPR010071"
                     /db_xref="UniProt/TrEMBL:Q54959"
                     /translation="MLVPDDARLPLTAAQLGIWFAQARDPHSPVFNTGEYVELTGGLD
                     ADRFEAALRATVARTDALRLRFHTDAEGTVTQSVADDVPFTLHRVDVSAEDDPAAAAD
                     AWMRDVIARPLDVTCGRPLFTQALFTLGPGRHRWFQHVHHLLLDGYGFRLVAERVAAA
                     YRAGAEAGDAEAGDTEPGIRDLLAAEADWRASDAYAQDRAHWLATPRHTPVTLAGTTA
                     PGAPGALRRTASAGPEVLAGLDAVARAAETGRQVAFPAAVAAYVQWLTGERTVTLGMP
                     VTARRTSAARRAPGTVANVPALHVEPRPAAPLAELVASVRDALEPLLRHQRYPYEEAR
                     RELGAVGERAYGPVVNVMPYGRALDFGGPAASVHPVSTGPVEDLKFSVYEEPDGSGAR
                     IDLEANAAVYGESETDAHLARFVRFLGAFAEAGPDTPLGRIHAAAPAAPAAAAPTTGG
                     PALHTLFAARAAATPHAVAVSHDGDHLTYAQLDARANRLAHRLVEQGAGPERLVALTL
                     PRSLDLVVALLAVLKSGAAYLPLDPDYPADRLAWMLEDAAPVLVLDPQAMTEDLAGYP
                     DTAPRTAVDGAHPAYVIYTSGSTGRPKGVVIPHSNVVRLFTSTDHWFGFGPDDVWTLF
                     HSYAFDFSVWEIWGALLHGGRLVVVPYHVSRSPGDFLDLLAREKVTVLNQTPTAFHQL
                     DAADRARTAAPELALRYVVFGGEALDVARLADWYARRGTAARLVNMYGITETTVHVTH
                     APLGPGHAVPGTPSLLGGPIPDLTPRVLDAALRPVPPGFTGELYVAGAGLARGYLNRP
                     ALTAQRFPADPYGAPGTRMYRTGDLVRHLDDGTYAYLGRGDDQVKIRGFRIELGEIEN
                     VLATHPGVAQAAAVVREDRHGDLRLAAYAVPTPGTEPDVAGWRALLAGQLPAHSLPAS
                     FTLLAALPLTANGKLDRAALPAPQAPAPETADEPAPAADPGEAALRRIFADTLGTPRV
                     GADSDFFTLGGHSLLAGRLIAAVRDTLGAEITIQDLFEAPTPAGLAARVRTAATGRPP
                     LVPRDHGDGPRPTSYAQHRMWVLREVEGAAPTYNMPLALRLSGTLNRTALRAALYDVV
                     VRHESLRTVYPRGEHGVTPRLLPPEQARPWVIESRVDDEQLPAALTAAAREPFNLDCQ
                     LPIRAHLFTLPGGDSVLLLLMHHVAGDGWSLRPLLGDLSTAYAARLAARAPDFPPLPV
                     SYADFAVWQRELLGDEHDPDSLAARQLAHWESTLAGLPDQLELPADFPRPPVAEHRGG
                     SVPLAVPPQLHARLDTFARAHGATCHMVLQAALAALLTRLGAGTDIPIGCPTASRPDS
                     ALDDLVGFFVNPVVLRADTSGDPTFTELLARVRATALAAYAHQDVPFDRVVERLQPER
                     SAARHPLFQVVLSYQDFEARAHLKGLDVRREDVDFGVAKFDLTFNLVEHRGDGEEFAG
                     IDGDLEYDTALYERATAEALTGRLLRLLDAVVTDPQAPVGSHDLLEEAEHARLAAFND
                     TARPVPRAGLAELFTAQARRTPDAVAVVGDDGTELTYAELDARSGALARRLAAAGAGR
                     ESAVAVLQQRSVHLVVSLLAVLKAGGFYVPLNTRYPADRMRLITADVKARVLLTDRAT
                     DTAHECRSWAGDADVLVVDELPTGGEPLPVVPAHPHQLAYVMYTSGSTGNPKGVAITH
                     GDVAALAADTCWRTGNQQRVLLHSPYSFDTSQYELWVPLLSGGTVVVAPPGDLDTRAL
                     ERAIVDGGVTGMWLTSGLFNLLAEESPHCFTGVREVWTGGDVVSPAAVARVLAASPST
                     MVADGYGPTETTTFATHHFMRAPWQQETTVPIGTPLDNTTCHVLDDRLRPVPPGVTGE
                     LYIGGSGLARGYLNRPELTAERFVADPYAPGGRMYRTGDLVRLRRDGVLTFLGRADHQ
                     VKVRGFRIELGEIESVLGAHPGIAQSAVLVREDRPGDKRLVAYVVPAPGATADATVLR
                     RHVAAALPDYMVPAAFVPLDALPLTGNGKLDRRALPVPDFGAADSAGRAPRTPREELL
                     AVLFAEVLGIEKVSIDDSFFDRGGDSIMVIQLVGRARRAGLLLNAAQIFEHKTVAQLA
                     PVAETLADTTREEPAAVAATGDVPLTPIMHWLRERGGPVDAFSQTMAVTVPAGLDRER
                     LVAALQTLIDHHDALRAVFSRDDADWTMRIEAPGTVDAADWVRHVDAAGLDEQALAAL
                     TTAEAAAARERLAPFAGRGGQAVWLDAGPHRPGRLLLTLHHLVVDGVSWRILLPDLKA
                     AWEATAPGSAAPALPPVTTSFRAWAHHLEQQAGCPERAAELALWTGMLAAGSPPVTAQ
                     PLDGRLDTAATARGLTLTLDTTTTAALLTTAPAAFNGRVDDILLTAFALAVTDWRRRR
                     GLGDSDAVLTDVEGHGREPGTSGAELSRTVGWFTSLYPVRLALGTLDAADALAGGPAA
                     GTAVKRIKEALRQIPDGGIGHGLLRHLDPRTGRSLAALAPYGTPPLGFNYLGRFPAEG
                     ADGLWSAAPDAEAIGDGTDPRLPLAHAVEVNALTLDEPAGPRLKATWTFAGRLLDEPA
                     VRDLGETWFRALNALVTHAERPEAGGITPSDLSLVQLSQDQIDQLEAAWKVSP"
     gene            7887. .22433
                     /gene="snbDE"
     CDS             7887. .22433
                     /gene="snbDE"
                     /codon_start=1
                     /transl_table=11
                     /product="Pristinamycin I synthase 3 and 4"
                     /protein_id="CAA72312.1"
                     /db_xref="GI:2052249"
                     /db_xref="GOA:O07944"
                     /db_xref="InterPro:IPR000051"
                     /db_xref="InterPro:IPR000379"
                     /db_xref="InterPro:IPR000873"
                     /db_xref="InterPro:IPR001031"
```

/db_xref="InterPro:IPR001242"
/db_xref="InterPro:IPR001601"
/db_xref="InterPro:IPR006162"
/db_xref="InterPro:IPR006163"
/db_xref="InterPro:IPR010071"
/db_xref="UniProt/TrEMBL:O07944"
/translation="MTRSGLADVLPLSPLQEGLLFQVTYGAETHGPDVYTVQMVFELR
GPLDEDGLRAAAAALLRRHPNLRAGFWQQGVERPVQPVNEVPLPWHTRDFTALGESD
RERAVEAYVTADRAERFDPGAPPLIRFALLGLAADHHKLVLTTHHLLLDGWSMPLLVR
ELFTLYGQRGDDAGMPPVTPYRAYLAHLAGRDDDAARAAWRTALADLEEPSLVAGAGA
GRGAADGSALPGQIWHEIDAATTAGLTALARSCNITLQHPGAKRLGTAARPQLGRDDV
VFGATVAHRPPEIPGIESTIGMFINTLPVRVRVRPAETLGDLLGRVQREQAALIEHRH
LSLTDIRSTTGTGELFDTVVVFENYPLDPAVLRAEARGLRLAGFEVSDATHYPLSLLA
IPGESLRFRLDHRGDVLDEAGARLLLERLDTLLTDIAEHGADLPVGRLDLLSAAERHQ
VLEEFNDTGLPAEDATLAALFEAQAARTPDTTALLVGGRSLTYAELNARANRLARHLV
TLGAGPEQIVAVKLERSLDLYVALLAVLKTGAAYLPVDTAYPAERIAFMMDDARPAVV
LTGEDTGQDLSGYDDTDLTDADRTAPLLPAHPAYVIYTSGSTGTPKAVVMPGAALVNL
LAWHRREIPGEAGAPVAQFTTIGFDVAAQEILATWLHGKTLAVPSQEVRRSAEQLAAW
LDEQHVSELYAPNLVIEALAEAAAEAGRTLPALRHIAQAGEALTLTRTVREFAAAVPG
RQLHNHYGPAETHVMTGTALPEDPAAWSEHAPLGRPVSGARVYVLDSALRPVAPGVTG
ELYLAGAGVSRGYLNRPVLTAERFVADPYAPSPGARMYRTGDLGRWNTRGELEFAGRA
DHQVKIRGFRIEPGEIEAALTDLPAVARAAVVVREDRPGDKRLVAYAVPAGEGLDAAA
VRSRLARTLPDFMVPAAIVALDALPLTPNGKLDRAALPAPQSTTAVQRTPRSPREEIL
TTLFAAVLKLPRVGIDDNFFDLGGHSLLATRLISRIRTVLGAEITLRDLFEAPTVAGL
GERLDQASGARPALTPAERPERLPLSHAQRRLWFLGRLDGPNSTYNIPLALRLSGTLD
TAALRAALADLVARHETLRTRYPSHDGEPYQHIVAPADATPALTVVHTEEHRLAEQLA
EATARPYDLTTELPLRATLFRLGEDEHVLLLLLHHIAGDGWSLAPLTRDLAAAYAARR
DGHAPGFAPLPVQYADYTLWQRGLLGDADDPTSLQAAQLAHWRQALEGRPAHLELPCD
HPRPAVATHRGATVPFHIDAGLHEKLTALAKACDSSLFMVLQAAFAALLTRHGAGTDI
PVGSPVAGRTDDALDDLVGFFVNTLVLRTDTSGDPTFRELVARVRQFDLAAYTHQDMP
FEKLVEEVNPERSLARNPLFQVVLALQSMPAADLALPGLDVAAEPVRVGFAKFDLGLA
VIEEHTADGTRAGIRGDWEYSTELFERGTVEALGSRLVRLLDAVADDPERTIGSVDLL
DATERHRLLVGLNTTDGPLPEERTLTALFEQQAAATPDATALVMGDRSLTYAELDARA
NRLARHLVALGAGPEQIVALQLPRSLDLVTAVLAVWKSGAAYLPVDPDYPPERIAHMR
ADARPALVVDAIPDDTTLAAYADSRLTDADRSAPLLPAHPAYVIYTSGSTGAPKGVVV
AHRSLAATVPAQAAAFGLGTHSRVLNFASISFDAAVWELTSALFTGAGLVLADADDLL
PGPSLARLVHDRHITLIALPPSALPALPDGALPPGTDLIVAGDATAPDQAARFAPGRR
MVNAYGLTETTVCATMSEPATGDGAPPIGRPVAHARVYVLDERLRPVPPGVTGEMYVS
GAGVARGYLHRPALTAQRFVPDPYALLFGETGTRMYRTGDLARLDADGRLHFAGRADQ
QVKIRGFRIEPGEIETVLTAHPAVAAGAVIARED TPGDKQLVAYLTRDTTHHAAPDQD
LGTGQVDRWQQTYDSLYEAEPGRAFGEDFSGWNSSYTAEEIPLEEMREWRAATVDRVL
ALRPRRVLEIGCGTGLILSQVAPHTEEYRGTDLSRSVVARLAAHLAGRPDLADKVTVD
ARAAHETDDLPKGRFDTIVLNSVAQYFPDARYLAGILHRAAELLAPGGTIFLGDIRNL
RTLHAFRTAVELRRAGPFADPAAVRRAVEQSLATEKELLLDPDFFTDLVDRDPHLTRA
DITLRTAAHHNEMSRHRYDVTLHRGGDSPAPAETVLRWGTDLDTAGELARVLSGPRPV
RVTGIPNARLTGETRAAQALERGSAADACARLEAPAPAGALDPHDVYALGAVCTWGHD
DDTFDAYAGDGTAYRPRTGRQGPLANDPARAEQDAALVGELRDLVAQRLPAHMAPAAY
VLLDRLPLSANGKLDRDALPAPDRGEDAAGQAPRTRREEILSTLFAEVLGLPAVGIDR
SFFDLGGHSLLATRLLSRIRSVLGAELAVRDLFQAPTVAALAERVDGAQEARPPLTAR
TRPAEVPLSFAQYRLWFLHRMEGPGPTYNIPMSLRLTGTLDTAALRAALADVTARHEA
LRTVYPESGGVPRQHVLHDVRPHFETADTTAQTLEADLTRAARRGFDLARELPLRATL
FRLAEDEHVLLLLMHHIAGDGWSWPPLARDLGDAYAARLAGRAPDLPPLPVQYADYAL
WQRDLLGDEDDPTSRHARQLAYWAGALDDLPEELQLPCDRPRPAVATHRGDTVPFAVD
PDLHQRLNALATEGRASLFMVLQAAFAALLTRHGAGTDIPLGSPIAGRTDDALEDLVG
FFVNTLVLRTDTSGDPTFRELLQRVRESDLAAYAHQDVPFEKLVEKLRPERSLSRHPL
FQVMLAFQNSGEARLALPGLDARALPVGVGVAKFDLHLSMVELRTGDGRPGGIQAALE
YATDLFDRTTAQALVDRLLRLLAAVADDPGLTVGRVELLDPAERHRVLQDWNTTPGGD
GAARTPVTELLERQAARTPDAVALVHDDGRLTYAELHARANRLARHLITLGAGPEQIV
ALRMPRSLDLYVALLAVLKTGAAYLPVDISYPAERIAFMIEDARPVTVLDRLPDDLGA
YRDTDLTDADRTAPLRPEHPAYVIHTSGSTGTPKAVVMPHAGLANLLTWHARRFPGGT
GVRTAQFTAIGFDFSVQEILSPLVMGKTLAVPSEEVRHSAELLAGWLETQQINELFAP
NLVIEALAEAAAEAGRTLPDLTDILQGGEALAPTERVRAFTAAVPGRRLHNVYGPAET
HAVTTHTLAADPAHWPPSAPIGRPVDHDRVYVLDSALRPVPPGVTGELYLAGAGVARG
YLHRPRTTAERFVADPYAPRPGARMYRTGDLGRWNTRGELEFAGRADHQVKIRGFRVE
PGEVEAALTAHPRITQAAVLAHGDRLVAYVVTTADRDLTGVREHLAARLPDFMVPTAY
VRLEALPLTVNGKLDRAALPAPDAPAAGRAPRTVREEILCGLFAEVLGTERAGIDDNF
FDLGGHSLLATRLISRIRTALGAEITLRDLFEAPTVAGLGERLDQASGARPALTPAER
PERLPLSYAQRRLWFLGRLTGGDPGYNMPVGLRLSGTLDTAALTAALADLVARHETLR
TVFPEGDDGSPHQLVLDPGAARPALHPRDTTQDDLATELAAAARHCFDLTTELPLRAT
LFRLAEDEHVLLLLLHHIAGDGWSLAPLARDLVTAYAARRDGHAPDFAPLPVQYADYT
LWQHALLGDENDPDSTISRQSAHWKDALAGLPEEIQLPADRPRPATPAHRGEAVFRTL
DADLHRELLHLARTGGASLFMVLQAAVAAVLTRHGAGTDIPLGSPIAGRTDDALDDLV
GFFVNTLVLRTDTSGNPTFRELLDRVREFDLAAYAHQDVPFEKLVEDAAPGRSLSRHP
LFQVLLALQNTPEAVLELPGLTARPEVVTLGAAKFDLTFNLAERHGPDGEPHGIDAVL
EYSTDLFDRTTAEALTDRLITFLRAVAADPGRTVADVPILDADDHARVLTGWNDTDHP
RGRRTADACLPRRIAEQAARTPHAVAVTEAGGTLITYSELDARANRLARHLIRRGVTA
ETRVAVLAERSAQLVVTTLAILKAGGVYVPLHTGYPVDRMRHVLADTEAALLLTDTHH
AATAARLDTPALTVDEDTTAGEQDTTAPDVTVRPDQLAYIMFTSGSTGTPKGIGITHR
DAIALAADRCWDLDTGSRVLMHSPYAFDISTFELFAPLLAGGRIVVAPRGDIDAAVLQ
RTLAAHGVTSLLLTAGLLGVIADEAPEVFTGVKDVWTGGDVVSPTAVRRVLEACPGTV
VKTLYGPTETTLGCTWLPFTDPRRIPPAVPIGRPLDNTRAYVLDERLRPVPPGVTGEL
YIAGAGLARGYWDQSARTAERFTADPHAHLFGDTGGRMYRTGDLARRDADGVLHFCGR
ADQQVKIRGFRIEPGEIETALAAHPDVTRAAVVARPGRAGDKVLVAYLVTAPDAGDTT
AEQLRAGLETQLPDYMVPAAFVALPALPVTPNGKLDRDALPEPDWGGGAGRPPRGPRE
EILCGLFAEVLGAPRVGTDDNFFELGGHSMLATRLVGRVKTVLGADIGVRTLFEAPTV
AALAARIDGADTGHDPFGVVLPLRTGGSRAPLFCVHPAGGFGWVYSALLRHTDREQPL
YALQARGLARPEPLPDDIDAMARDYAEQIRKTVPEGPYEILGWSFGGLVAHAVAARLQ
AEGAEVSLLAVLDGYPDAYDGTEHEVGEEQVLAILLNAAGVDRAQAFGDAPLQRAAVL
EKLRDSGSALGNLDDDAVGRMVTVFLNNTRLIQNFRPRRFTGDMVFFAAAAGHHDPAL
TPGNWRPYVTGRIEEHHLDTDHAGLARPEALGAVARTLAQRDTPQAGH"

ORIGIN

Query Match 10.8%; Score 108.2; DB 1; Length 22449;
Best Local Similarity 46.6%; Pred. No. 0.43;
Matches 449; Conservative 0; Mismatches 508; Indels 6; Gaps 3;



```
Db      21876 TACGAGATCCTCGGCTGGTCCTTCGGCGGCCTCGTCGCCCACGCCGTCGCCGCCCGCCTG 21935

Qy        792 CTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCG 851
              |   | | || |     ||   || ||||| | | ||||||  ||| ||||  | |
Db      21936 CAGGCCGAGGGCGCCGAAGTGAGCCTGCTGGCCGTCCTCGACGGCTACCCCGACGCCTAC 21995

Qy        852 GGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCC 911
              | |  | |  || |   |  |||||    || | | ||||| || |||| |||| |||
Db      21996 GACGGCACCGAGCACGAGGTCGGCGA---GGAACAGGTCCTGGCGATCCTCCTCAACGCC 22052

Qy        912 AACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCGACGACCTCGCC 971
                | | ||  | |  | |   |  ||| ||| | |  || | |  | || | |   |
Db      22053 GCCGGCGTCGACCGGGCCCAGGCCTTCGGCGACGCCCCCCTCCAACGGGCCGCCGTGCTC 22112

Qy        972 GGG 974
              | |
Db      22113 GAG 22115
```

RESULT 9
SPSNBCGEN
LOCUS SPSNBCGEN 22449 bp DNA linear BCT 18-APR-2005
DEFINITION S.pristinaespiralis snbC and snbDE genes.
ACCESSION X98690
VERSION X98690.1 GI:1906377
KEYWORDS Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbC gene; snbDE gene.
SOURCE Streptomyces pristinaespiralis
ORGANISM Streptomyces pristinaespiralis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS de Crecy-Lagard,V., Blanc,V., Gil,P., Naudin,L., Lorenzon,S., Famechon,A., Bamas-Jacques,N., Crouzet,J. and Thibaut,D.
TITLE Pristinamycin I biosynthesis in Streptomyces pristinaespiralis: molecular characterization of the first two structural peptide synthetase genes
JOURNAL J. Bacteriol. 179 (3), 705-713 (1997)
PUBMED 9006024
REFERENCE 2
AUTHORS de Crecy-Lagard,V.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1996) V. de Crecy-Lagard, Institut Pasteur, 25 Rue de Dr. Roux, Paris Cedex 75724, FRANCE
REMARK Revised by [3]
REFERENCE 3 (bases 1 to 22449)
AUTHORS de Crecy-Lagard,V.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1997) V. de Crecy-Lagard, Institut Pasteur, 25 Rue de Dr. Roux, Paris Cedex 75724, FRANCE
COMMENT On Mar 25, 1997 this sequence version replaced gi:1483602.
FEATURES Location/Qualifiers

```
     source          1. .22449
                     /organism="Streptomyces pristinaespiralis"
                     /mol_type="genomic DNA"
                     /strain="SP92"
                     /db_xref="taxon:38300"
                     /clone="pIBV3 cosmid"
     gene            115. .7890
                     /gene="snbC"
     CDS             115. .7890
                     /gene="snbC"
                     /codon_start=1
                     /transl_table=11
                     /product="Pristinamycin I synthase 2"
                     /protein_id="CAA67248.1"
                     /db_xref="GI:1483603"
                     /db_xref="GOA:Q54959"
                     /db_xref="InterPro:IPR000873"
                     /db_xref="InterPro:IPR001242"
                     /db_xref="InterPro:IPR006163"
                     /db_xref="InterPro:IPR010060"
                     /db_xref="InterPro:IPR010071"
                     /db_xref="UniProt/TrEMBL:Q54959"
                     /translation="MLVPDDARLPLTAAQLGIWFAQARDPHSPVFNTGEYVELTGGLD
                     ADRFEAALRATVARTDALRLRFHTDAEGTVTQSVADDVPFTLHRVDVSAEDDPAAAAD
                     AWMRDVIARPLDVTCGRPLFTQALFTLGPGRHRWFQHVHHLLLDGYGFRLVAERVAAA
                     YRAGAEAGDAEAGDTEPGIRDLLAAEADWRASDAYAQDRAHWLATPRHTPVTLAGTTA
                     PGAPGALRRTASAGPEVLAGLDAVARAAETGRQVAFPAAVAAYVQWLTGERTVTLGMP
                     VTARRTSAARRAPGTVANVPALHVEPRPAAPLAELVASVRDALEPLLRHQRYPYEEAR
                     RELGAVGERAYGPVVNVMPYGRALDFGGPAASVHPVSTGPVEDLKFSVYEEPDGSGAR
                     IDLEANAAVYGESETDAHLARFVRFLGAFAEAGPDTPLGRIHAAAPAAPAAAAPTTGG
                     PALHTLFAARAAATPHAVAVSHDGDHLTYAQLDARANRLAHRLVEQGAGPERLVALTL
                     PRSLDLVVALLAVLKSGAAYLPLDPDYPADRLAWMLEDAAPVLVLDPQAMTEDLAGYP
                     DTAPRTAVDGAHPAYVIYTSGSTGRPKGVVIPHSNVVRLFTSTDHWFGFGPDDVWTLF
                     HSYAFDFSVWEIWGALLHGGRLVVVPYHVSRSPGDFLDLLARAEKVTVLNQTPTAFHQL
                     DAADRARTAAPELALRYVVFGGEALDVARLADWYARRGTAARLVNMYGITETTVHVTH
                     APLGPGHAVPGTPSLLGGPIPDLTPRVLDAALRPVPPGFTGELYVAGAGLARGYLNRP
                     ALTAQRFPADPYGAPGTRMYRTGDLVRHLDDGTYAYLGRGDDQVKIRGFRIELGEIEN
                     VLATHPGVAQAAAVVREDRHGDLRLAAYAVPTPGTEPDVAGWRALLAGQLPAHSLPAS
                     FTLLAALPLTANGKLDRAALPAPQAPAPETADEPAPAADPGEAALRRIFADTLGTPRV
                     GADSDFFTLGGHSLLAGRLIAAVRDTLGAEITIQDLFEAPTPAGLAARVRTAATGRPP
                     LVPRDHGDGPRPTSYAQHRMWVLREVEGAAPTYNMPLALRLSGTLNRTALRAALYDVV
                     VRHESLRTVYPRGEHGVTPRLLPPEQARPWVIESRVDDEQLPAALTAAAREPFNLDCQ
                     LPIRAHLFTLPGGDSVLLLLMHHVAGDGWSLRPLLGDLSTAYAARLAARAPDFPPLPV
                     SYADFAVWQRELLGDEHDPDSLAARQLAHWESTLAGLPDQLELPADFPRPPVAEHRGG
                     SVPLAVPPQLHARLDTFARAHGATCHMVLQAALAALLTRLGAGTDIPIGCPTASRPDS
                     ALDDLVGFFVNPVVLRADTSGDPTFTELLARVRATALAAYAHQDVPFDRVVERLQPER
                     SAARHPLFQVVLSYQDFEARAHLKGLDVRREDVDFGVAKFDLTFNLVEHRGDGEEFAG
                     IDGDLEYDTALYERATAEALTGRLLRLLDAVVTDPQAPVGSHDLLEEAEHARLAAFND
                     TARPVPRAGLAELFTAQARRTPDAVAVVGDDGTELTYAELDARSGALARRLAAAGAGR
                     ESAVAVLQQRSVHLVVSLLAVLKAGGFYVPLNTRYPADRMRLITADVKARVLLTDRAT
                     DTAHECRSWAGDADVLVVDELPTGGEPLPVVPAHPHQLAYVMYTSGSTGNPKGVAITH
                     GDVAALAADTCWRTGNQQRVLLHSPYSFDTSQYELWVPLLSGGTVVVAPPGDLDTRAL
                     ERAIVDGGVTGMWLTSGLFNLLAEESPHCFTGVREVWTGGDVVSPAAVARVLAASPST
                     MVADGYGPTETTTFATHHFMRAPWQQETTVPIGTPLDNTTCHVLDDRLRPVPPGVTGE
                     LYIGGSGLARGYLNRPELTAERFVADPYAPGGRMYRTGDLVRLRRDGVLTFLGRADHQ
                     VKVRGFRIELGEIESVLGAHPGIAQSAVLVREDRPGDKRLVAYVVPAPGATADATVLR
                     RHVAAALPDYMVPAAFVPLDALPLTGNGKLDRRALPVPDFGAADSAGRAPRTPREELL
                     AVLFAEVLGIEKVSIDDSFFDRGGDSIMVIQLVGRARRAGLLLNAAQIFEHKTVAQLA
                     PVAETLADTTREEPAAVAATGDVPLTPIMHWLRERGGPVDAFSQTMAVTVPAGLDRER
                     LVAALQTLIDHHDALRAVFSRDDADWTMRIEAPGTVDAADWVRHVDAAGLDEQALAAL
                     TTAEAAAARERLAPFAGRGGQAVWLDAGPHRPGRLLLTLHHLVVDGVSWRILLPDLKA
                     AWEATAPGSAAPALPPVTTSFRAWAHHLEQQAGCPERAAELALWTGMLAAGSPPVTAQ
                     PLDGRLDTAATARGLTLTLDTTTTAALLTTAPAAFNGRVDDILLTAFALAVTDWRRRR
                     GLGDSDAVLTDVEGHGREPGTSGAELSRTVGWFTSLYPVRLALGTLDAADALAGGPAA
                     GTAVKRIKEALRQIPDGGIGHGLLRHLDPRTGRSLAALAPYGTPPLGFNYLGRFPAEG
                     ADGLWSAAPDAEAIGDGTDPRLPLAHAVEVNALTLDEPAGPRLKATWTFAGRLLDEPA
                     VRDLGETWFRALNALVTHAERPEAGGITPSDLSLVQLSQDQIDQLEAAAWKVSP"
gene                 7887. .22433
                     /gene="snbDE"
CDS                  7887. .22433
                     /gene="snbDE"
                     /codon_start=1
                     /transl_table=11
                     /product="Pristinamycin I synthase 3 and 4"
                     /protein_id="CAA67249.1"
                     /db_xref="GI:1906378"
                     /db_xref="GOA:O07944"
                     /db_xref="InterPro:IPR000051"
                     /db_xref="InterPro:IPR000379"
                     /db_xref="InterPro:IPR000873"
                     /db_xref="InterPro:IPR001031"
                     /db_xref="InterPro:IPR001242"
                     /db_xref="InterPro:IPR001601"
                     /db_xref="InterPro:IPR006162"
                     /db_xref="InterPro:IPR006163"
                     /db_xref="InterPro:IPR010071"
                     /db_xref="UniProt/TrEMBL:O07944"
                     /translation="MTRSGLADVLPLSPLQEGLLFQVTYGAETHGPDVYTVQMVFELR
                     GPLDEDGLRAAAAALLRRHPNLRAGFWQQGVERPVQFVPNEVPLPWHTRDFTALGESD
                     RERAVEAYVTADRAERFDPGAPPLIRFALLGLAADHHKLVLTTHHLLLDGWSMPLLVR
                     ELFTLYGQRGDDAGMPPVTPYRAYLAHLAGRDDDAARAAWRTALADLEEPSLVAGAGA
                     GRGAADGSALPGQIWHEIDAATTAGLTALARSCNITLQHPGAKRLGTAARPQLGRDDV
                     VFGATVAHRPPEIPGIESTIGMFINTLPVRVRVRPAETLGDLLGRVQREQAALIEHRH
                     LSLTDIRSTTGTGELFDTVVVFENYPLDPAVLRAEARGLRLAGFEVSDATHYPLSLLA
```

IPGESLRFRLDHRGDVLDEAGARLLLERLDTLLTDIAEHGADLPVGRLDLLSAAERHQ
VLEEFNDTGLPAEDATLAALFEAQAARTPDTTALLVGGRSLTYAELNARANRLARHLV
TLGAGPEQIVAVKLERSLDLYVALLAVLKTGAAYLPVDTAYPAERIAFMMDDARPAVV
LTGEDTGQDLSGYDDTDLTDADRTAPLLPAHPAYVIYTSGSTGTPKAVVMPGAALVNL
LAWHRREIPGEAGAPVAQFTTIGFDVAAQEILATWLHGKTLAVPSQEVRRSAEQLAAW
LDEQHVSELYAPNLVIEALAEAAAEAGRTLPALRHIAQAGEALTLTRTVREFAAAVPG
RQLHNHYGPAETHVMTGTALPEDPAAWSEHAPLGRPVSGARVYVLDSALRPVAPGVTG
ELYLAGAGVSRGYLNRPVLTAERFVADPYAPSPGARMYRTGDLGRWNTRGELEFAGRA
DHQVKIRGFRIEPGEIEAALTDLPAVARAAVVVREDRPGDKRLVAYAVPAGEGLDAAA
VRSRLARTLPDFMVPAAIVALDALPLTPNGKLDRAALPAPQSTTAVQRTPRSPREEIL
TTLFAAVLKLPRVGIDDNFFDLGGHSLLATRLISRIRTVLGAEITLRDLFEAPTVAGL
GERLDQASGARPALTPAERPERLPLSHAQRRLWFLGRLDGPNSTYNIPLALRLSGTLD
TAALRAALADLVARHETLRTRYPSHDGEPYQHIVAPADATPALTVVHTEEHRLAEQLA
EATARPYDLTTELPLRATLFRLGEDEHVLLLLLHHIAGDGWSLAPLTRDLAAAYAARR
DGHAPGFAPLPVQYADYTLWQRGLLGDADDPTSLQAAQLAHWRQALEGRPAHLELPCD
HPRPAVATHRGATVPFHIDAGLHEKLTALAKACDSSLFMVLQAAFAALLTRHGAGTDI
PVGSPVAGRTDDALDDLVGFFVNTLVLRTDTSGDPTFRELVARVRQFDLAAYTHQDMP
FEKLVEEVNPERSLARNPLFQVVLALQSMPAADLALPGLDVAAEPVRVGFAKFDLGLA
VIEEHTADGTRAGIRGDWEYSTELFERGTVEALGSRLVRLLDAVADDPERTIGSVDLL
DATERHRLLVGLNTTDGPLPEERTLTALFEQQAAATPDATALVMGDRSLTYAELDARA
NRLARHLVALGAGPEQIVALQLPRSLDLVTAVLAVWKSGAAYLPVDPDYPPERIAHMR
ADARPALVVDAIPDDTTLAAYADSRLTDADRSAPLLPAHPAYVIYTSGSTGAPKGVVV
AHRSLAATVPAQAAAFGLGTHSRVLNFASISFDAAVWELTSALFTGAGLVLADADDLL
PGPSLARLVHDRHITLIALPPSALPALPDGALPPGTDLIVAGDATAPDQAARFAPGRR
MVNAYGLTETTVCATMSEPATGDGAPPIGRPVAHARVYVLDERLRPVPPGVTGEMYVS
GAGVARGYLHRPALTAQRFVPDPYALLFGETGTRMYRTGDLARLDADGRLHFAGRADQ
QVKIRGFRIEPGEIETVLTAHPAVAAGAVIAREDTPGDKQLVAYLTRDTTHHAAPDQD
LGTGQVDRWQQTYDSLYEAEPGRAFGEDFSGWNSSYTAEEIPLEEMREWRAATVDRVL
ALRPRRVLEIGCGTGLILSQVAPHTEEYRGTDLSRSVVARLAAHLAGRPDLADKVTVD
ARAAHETDDLPKGRFDTIVLNSVAQYFPDARYLAGILHRAAELLAPGGTIFLGDIRNL
RTLHAFRTAVELRRAGPFADPAAVRRAVEQSLATEKELLLDPDFFTDLVDRDPHLTRA
DITLRTAAHHNEMSRHRYDVTLHRGGDSPAPAETVLRWGTDLDTAGELARVLSGPRPV
RVTGIPNARLTGETRAAQALERGSAADACARLEAPAPAGALDPHDVYALGAVCTWGHD
DDTFDAYAGDGTAYRPRTGRQGPLANDPARAEQDAALVGELRDLVAQRLPAHMAPAAY
VLLDRLPLSANGKLDRDALPAPDRGEDAAGQAPRTRREEILSTLFAEVLGLPAVGIDR
SFFDLGGHSLLATRLLSRIRSVLGAELAVRDLFQAPTVAALAERVDGAQEARPPLTAR
TRPAEVPLSFAQYRLWFLHRMEGPGPTYNIPMSLRLTGTLDTAALRAALADVTARHEA
LRTVYPESGGVPRQHVLHDVRPHFETADTTAQTLEADLTRAARRGFDLARELPLRATL
FRLAEDEHVLLLLMHHIAGDGWSWPPLARDLGDAYAARLAGRAPDLPPLPVQYADYAL
WQRDLLGDEDDPTSRHARQLAYWAGALDDLPEELQLPCDRPRPAVATHRGDTVPFAVD
PDLHQRLNALATEGRASLFMVLQAAFAALLTRHGAGTDIPLGSPIAGRTDDALEDLVG
FFVNTLVLRTDTSGDPTFRELLQRVRESDLAAYAHQDVPFEKLVEKLRPERSLSRHPL
FQVMLAFQNSGEARLALPGLDARALPVGVGVAKFDLHLSMVELRTGDGRPGGIQAALE
YATDLFDRTTAQALVDRLLRLLAAVADDPGLTVGRVELLDPAERHRVLQDWNTTPGGD
GAARTPVTELLERQAARTPDAVALVHDDGRLTYAELHARANRLARHLITLGAGPEQIV
ALRMPRSLDLYVALLAVLKTGAAYLPVDISYPAERIAFMIEDARPVTVLDRLPDDLGA
YRDTDLTDADRTAPLRPEHPAYVIHTSGSTGTPKAVVMPHAGLANLLTWHARRFPGGT
GVRTAQFTAIGFDFSVQEILSPLVMGKTLAVPSEEVRHSAELLAGWLETQQINELFAP
NLVIEALAEAAAEAGRTLPDLTDILQGGEALAPTERVRAFTAAVPGRRLHNVYGPAET
HAVTTHTLAADPAHWPPSAPIGRPVDHDRVYVLDSALRPVPPGVTGELYLAGAGVARG
YLHRPRTTAERFVADPYAPRPGARMYRTGDLGRWNTRGELEFAGRADHQVKIRGFRVE
PGEVEAALTAHPRITQAAVLAHGDRLVAYVVTTADRDLTGVREHLAARLPDFMVPTAY
VRLEALPLTVNGKLDRAALPAPDAPAAGRAPRTVREEILCGLFAEVLGTERAGIDDNF
FDLGGHSLLATRLISRIRTALGAEITLRDLFEAPTVAGLGERLDQASGARPALTPAER
PERLPLSYAQRRLWFLGRLTGGDPGYNMPVGLRLSGTLDTAALTAALADLVARHETLR
TVFPEGDDGSPHQLVLDPGAARPALHPRDTTQDDLATELAAAARHCFDLTTELPLRAT
LFRLAEDEHVLLLLLHHIAGDGWSLAPLARDLVTAYAARRDGHAPDFAPLPVQYADYT
LWQHALLGDENDPDSTISRQSAHWKDALAGLPEEIQLPADRPRPATPAHRGEAVFRTL
DADLHRELLHLARTGGASLFMVLQAAVAAVLTRHGAGTDIPLGSPIAGRTDDALDDLV
GFFVNTLVLRTDTSGNPTFRELLDRVREFDLAAYAHQDVPFEKLVEDAAPGRSLSRHP
LFQVLLALQNTPEAVLELPGLTARPEVVTLGAAKFDLTFNLAERHGPDGEPHGIDAVL
EYSTDLFDRTTAEALTDRLITFLRAVAADPGRTVADVPILDADDHARVLTGWNDTDHP
RGRRTADACLPRRIAEQAARTPHAVAVTEAGGTLITYSELDARANRLARHLIRRGVTA
ETRVAVLAERSAQLVVTTLAILKAGGVYVPLHTGYPVDRMRHVLADTEAALLLTDTHH
AATAARLDTPALTVDEDTTAGEQDTTAPDVTVRPDQLAYIMFTSGSTGTPKGIGITHR
DAIALAADRCWDLDTGSRVLMHSPYAFDISTFELFAPLLAGGRIVVAPRGDIDAAVLQ
RTLAAHGVTSLLLTAGLLGVIADEAPEVFTGVKDVWTGGDVVSPTAVRRVLEACPGTV
VKTLYGPTETTLGCTWLPFTDPRRIPPAVPIGRPLDNTRAYVLDERLRPVPPGVTGEL
YIAGAGLARGYWDQSARTAERFTADPHAHLFGDTGGRMYRTGDLARRDADGVLHFCGR
ADQQVKIRGFRIEPGEIETALAAHPDVTRAAVVARPGRAGDKVLVAYLVTAPDAGDTT
AEQLRAGLETQLPDYMVPAAFVALPALPVTPNGKLDRDALPEPDWGGGAGRPPRGPRE
EILCGLFAEVLGAPRVGTDDNFFELGGHSMLATRLVGRVKTVLGADIGVRTLFEAPTV
AALAARIDGADTGHDPFGVVLPLRTGGSRAPLFCVHPAGGFGWVYSALLRHTDREQPL

YALQARGLARPEPLPDDIDAMARDYAEQIRKTVPEGPYEILGWSFGGLVAHAVAARLQ
AEGAEVSLLAVLDGYPDAYDGTEHEVGEEQVLAILLNAAGVDRAQAFGDAPLQRAAVL

Query Match 10.8%; Score 108.2; DB 1; Length 22449;
Best Local Similarity 46.6%; Pred. No. 0.43;
Matches 449; Conservative 0; Mismatches 508; Indels 6; Gaps 3;

```
Qy       12 GTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGGCGCTCCGCAAC 71
           |  |  ||  |  ||  |     |  ||||  |  |||   |  |   |  ||   |||  ||  |
Db    21159 GCCGTCGTCGCCCGCCCCGGCCGGGCCGGCGACAAGGTCCTCGTCGCCTACCTCGTCACC 21218

Qy       72 TACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGG 131
           |  |  |     |  |  | ||||||      ||||  |  |   |   ||   |  |
Db    21219 GCCCCCGACGCGGGCGACACCACCGCCGAACAACTGCGCGCCGGCCTCGAGACCCAACTG 21278

Qy      132 TCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCAC 191
           |  |     |  ||  |||||||  |  |   |   ||      |  |||||  ||  ||
Db    21279 CCCGACTACATGGTCCCCGCCGCGTTCGTCGCGCTGCCCGCCCTGCCCGTCACGCCCAAC 21338

Qy      192 ATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCC 251
              |   |  ||     |  ||  ||  |  ||  |     |  |  ||  |||||   ||||
Db    21339 GGCAAACTCGACC--GCGACGCACTGCCCGAACCCGACTGGGGCGGCGGCGCCGGACGCC 21396

Qy      252 GCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCG 311
             ||  ||  |  |       |   ||   |||||||   |  |   |  |  |  |  |
Db    21397 CGCCCCGCGGCCCCCGCGAGGAAATCCTGTGCGGCCTGTTCGCCGAGGTCCTCGGCGCCC 21456

Qy      312 GCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCC 371
             ||  |  ||||  |  |  |  |  |  |  |  |  ||  |  |  |  |  ||  | |||
Db    21457 CCCGGGTCGGCACCGACGACAACTTCTTCGAACTCGGCGGCCACTCCATGCTCGCCACCC 21516

Qy      372 GCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCG 431
             ||  |  ||  ||  ||  ||  |  |  |  ||||||      |||  ||  ||  |  ||
Db    21517 GCCTCGTCGGCCGCGTCAAGACCGTCCTCGGCGCCGACATCGGCGTACGCACCCTCTTCG 21576

Qy      432 CGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCC 491
             |     |   |  |  ||||  |  ||  |     |  |||  |  |  |  |  ||
Db    21577 AGGCGCCGACCGTCGCCGCCCTCGCCGCCCGCATCGACGGCGCCGACACGGGCCACGACC 21636

Qy      492 AACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTG 551
              |  ||  ||  |  |  |     |  |  |  |||||  ||||  |  ||||||  ||
Db    21637 -CCTTCGGCGTGGTGCTGCCGCTGCGCACCGGCGGCAGCCGCGCCCCCCTGTTCTGCGTC 21695

Qy      552 CACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGAT 611
             |||  |||     ||  |||     |     ||   |  |||     |   ||  |   ||
Db    21696 CACCCCGCCGGCGGCTTCGGCTGGGTCTACTCCGCACTGCTGCGCCACACCGACCGCGAA 21755

Qy      612 TACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGC 671
             |  ||  |  ||||  |||||||  ||  |||  |  |  |  |  ||     |   |
Db    21756 CAGCCCCTGTACGCCCTCCAGGCACGCGGCCTGGCCCGCCCCGAGCCACTGCCCGACGAC 21815

Qy      672 ATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCT 731
             |||||  |  ||  ||   |  ||||  |     |  ||  |     ||   ||  ||
Db    21816 ATCGACGCCATGGCCCGCGACTACGCAGAACAGATCCGCAAAACGGTGCCCGAGGGCCCC 21875

Qy      732 TACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTC 791
             ||||    |   ||  |||||  |||||||  ||  |||||   ||    |  |||    |  ||
Db    21876 TACGAGATCCTCGGCTGGTCCTTCGGCGGCCTCGTCGCCCACGCCGTCGCCGCCCGCCTG 21935

Qy      792 CTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCG 851
             |   |  | ||  |    ||   ||  |||||  |  |  ||||||  |||  ||||   |  |
Db    21936 CAGGCCGAGGGCGCCGAAGTGAGCCTGCTGGCCGTCCTCGACGGCTACCCCGACGCCTAC 21995

Qy      852 GGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCC 911
             |  |  |  ||  |  |  |||||    ||  |  |  |||||  ||  ||||  ||||   |||
Db    21996 GACGGCACCGAGCACGAGGTCGGCGA---GGAACAGGTCCTGGCGATCCTCCTCAACGCC 22052

Qy      912 AACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCGACGACCTCGCC 971
             |  |  ||  |  |  |  |  |||  |||  |  |  ||  |  |  |  |  |  |  |
Db    22053 GCCGGCGTCGACCGGGCCCAGGCCTTCGGCGACGCCCCCCTCCAACGGGCCGCCGTGCTC 22112

Qy      972 GGG 974
             | |
```

```
Db      22113 GAG 22115


RESULT 10
AY787762
LOCUS       AY787762               127559 bp    DNA     linear   BCT 09-MAY-2005
DEFINITION  Streptomyces filamentosus strain NRRL 11379 daptomycin biosynthetic
            gene cluster, partial sequence.
ACCESSION   AY787762
VERSION     AY787762.1  GI:60650890
KEYWORDS    .
SOURCE      Streptomyces filamentosus (Streptomyces roseosporus)
  ORGANISM  Streptomyces filamentosus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 127559)
  AUTHORS   Miao,V., Coeffet-Legal,M.F., Brian,P., Brost,R., Penn,J.,
            Whiting,A., Martin,S., Ford,R., Parr,I., Bouchard,M., Silva,C.J.,
            Wrigley,S.K. and Baltz,R.H.
  TITLE     Daptomycin biosynthesis in Streptomyces roseosporus: cloning and
            analysis of the gene cluster and revision of peptide
            stereochemistry
  JOURNAL   Microbiology (Reading, Engl.) 151 (PT 5), 1507-1523 (2005)
   PUBMED   15870461
REFERENCE   2  (bases 1 to 127559)
  AUTHORS   Miao,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2004) Drug Discovery and Evaluation, Cubist
            Pharmaceuticals, Inc., 65 Hayden Av, Lexington, MA 02421, USA
FEATURES             Location/Qualifiers
     source          1. .127559
                     /organism="Streptomyces filamentosus"
                     /mol_type="genomic DNA"
                     /strain="NRRL 11379"
                     /db_xref="taxon:67294"
     CDS             complement(<1. .965)
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAX31516.1"
                     /db_xref="GI:60650891"
                     /translation="MDREAEAPLRAAPHATPAERAALGKAARREAPRSGHAEFSPSPR
                     RPDPLTVLEAQSADRVPELVPIRYARMTESPFRFYRGAAALMAADLAGTPVSGIRAQL
                     CGDAHLLNFRLLASPERNLLFDINDFDETLPGPWEWDVKRLAASLVIAGRANSFTLRE
                     RAGVVRATVRSYREAMARFAGMRNLDVWYARTDAERLRTVATEQLGGRGRRNVDRALG
                     KARSRDSLQAFGKLAEVVDGRLRIAADPPMVVPLTDLTPGVDRDAVFRQFGSMLAGYA
                     RSLPSDRRSLLEDFALVDVARKVVGVGSVGTRCWIVLLLGRDGGDP"
     CDS             complement(989. .1948)
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="AAX31517.1"
                     /db_xref="GI:60650892"
                     /translation="MIHIRAVSPPDLTDEVVGLLSADPCVLNLIVQRDAARRPDGDAI
                     ACDVLTGAANDVLHRLRAAHLDRRGSLVIEPVDMAFSGAATEGGQRELGPLSRAPVWE
                     QVEARIRSGGRYPPSFYLYLVIAGLIGSVGIVTNSQILIVGAMVVGPEYGAIVSVALG
                     IDRRHRSMVRSGLAALGVGLLLTIVVTFLFALLIRGFGLESEAFDRGLRPVSHLINTP
                     NFFSVAVATLAGIVGIVSLTEARTSALLGVFISVTTIPAAADIAVSTAYTSWSDVRGS
                     AIQLVVNILVLIVVGAFALKAQRAIWQRVRLRRDRERRIAEQA"
     CDS             2099. .2392
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAX31518.1"
                     /db_xref="GI:60650893"
                     /translation="MTRPGWDHEGVDTPDTPDAFPEPLPGADEAVREERATDDGTPEG
                     RRLVRCRLCGRPLTGADSRRAGLGPSCDAKLHPAPPDIRTRRHEVDQDPLPGT"
     CDS             complement(2405. .3277)
                     /codon_start=1
                     /transl_table=11
                     /product="putative acyl-coA thioesterase II"
                     /protein_id="AAX31519.1"
                     /db_xref="GI:60650894"
                     /translation="MTNPAERLVDLLDLERIEVNIFRGRSPEESLQRVFGGQVAGQAL
                     VAAGRTTDGERPVHSLHAYFLRPGRPGVPIVYQVERVRDGRSFTTRRVTAVQEGRTIF
                     NLTASFHRPEEAGFEHQLPPARIVPDPEELPTVAEEVREHLGALPEALERMARRQPFD
                     IRYVDRLRWTKDEIQDADPRSAVWMRAVGPLGDDPLVHTCALTYASDMTLLDAVRIPV
                     EPLWGPRGYDLASLDHAMWFHRPFRADEWFLYDQESPIATGGRGLARGRIYDRSGQLL
                     VSVVQEGLFRRLEQ"
     CDS             complement(3312. .5885)
                     /codon_start=1
                     /transl_table=11
                     /product="putative ATP-dependent helicase"
                     /protein_id="AAX31520.1"
                     /db_xref="GI:60650895"
                     /translation="MIFVPSAGSLIRAEDRQDGGVTLIDQLPQTADPDALFEAFSSWT
                     ESQGITMYPAQEEALIEVVSGANVILSTPTGSGKSLVAAGAHFTALAQDKVTFYTAPI
                     KALVSEKFFDLCKLFGTENVGMLTGDASVNADAPVICCTAEVLASIALRDGKYADIGQ
                     VVMDEFHFYAEPDRGWAWQIPLLELPQAQFVLMSATLGDVSMFEKDLTRRTGRPTSVV
                     RSATRPVPLSYEYRFTPITETLTELLDTRQSPVYIVHFTQAAAVERAQSLMSINMCTK
                     EEKEKIADLIGSFRFTTKFGQNLSRYVRHGIGVHHAGMLPKYRRLVEKLAQAGLLKVI
                     CGTDTLGVGVNVPIRTVLFTALTKYDGNRVRTLRAREFHQIAGRAGRAGFDTAGFVVA
                     QAPEHVIENEKALKKAGDDPKKKRKVVRKKAPEGFVAWSESTFDKLIQSEPEPLTSRF
                     RVTHTMLLAVIARPGNAFEAMRHLLEDNHEPRRAQLRHIRRAIAIYRSLLDGGVVEQL
                     DTPDAEGRIVRLTVDLQQDFALNQPLSTFALAAFDLLDAESPSYALDMVSVVESTLDD
                     PRQILAAQQNKARGEAVGQMKADGVEYEERMERLQEVTYPKPLSELLWHAYDVYRTSH
                     PWVNDHPVSPKSVIRDMYERAMTFTEFTSHYELARTEGIVLRYLASAYKALEHTIPDD
                     VKSEDLQDLISWLGEMVRQVDSSLLDEWEQLANPEVETAEQAQEKADEVKPVTANARA
                     FRVLVRNAMFRRVELAALDRAGALGELDGESGWDEDAWGEALDAYWDAHEEIGTGPDA
                     RGPKLLKIEEDPAHGLWRVWQAFADPAGDHDWGIKAEVDLAASDEEGRAVVRVTEVGQ
                     L"
     CDS             5963. .6754
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="AAX31521.1"
                     /db_xref="GI:60650896"
                     /translation="MMGPAHSLSGAAAWLGVGAAAAAAGHTMPWPVLVVGALICAGAA
                     LAPDLDHKSATISRAFGPVSKALCEIVDKLSYAVYKATKSAGDPRRTGGHRTLTHTWL
                     WAVLIGGGCSVAAITGGRWAVLVILFVHLVLAVEGLLWRAARVSSDVLVWLLGATSAW
                     ILAGVLDKPGYGADWLFDAPGQEYMWLGLPIVLGALVHDIGDALTVSGCPILWPIPIG
                     RKRWYPIGPPKAMRFRAGSWVEMKVLMPAFMVLGGVGGAAALNYI"
     CDS             6850. .8385
                     /codon_start=1
                     /transl_table=11
                     /product="putative DNA ligase"
                     /protein_id="AAX31522.1"
                     /db_xref="GI:60650897"
                     /translation="MLLAELAQVSLEVAATSARSKKVALLAGLFRDAGPEDVPVVIPY
                     LAGRLPQGRIGVGWRSLGDPVEPAAEPTLTVTGVDARLTALAAVSGPGSQARRKEHLR
                     ALFAAATEDEQRFLRALLTGEVRQGALDALAADALARAADAPPADVRRAVMLAGSLQE
                     VAGVLLADGPEALAAFRLTVGRPVQPMLAHTAASVGEALDKLGACAVEEKLDGIRVQV
                     HRDGDRIRAYTRTLDDITDRLPELTAAVAALPAGRFILDGEVIALGEDGRPRPFQETA
                     SRVGSRRDVAEAAAHVPVAPVFFDALLVDDEDLLDLPFTDRHAALARLLPEHLRVRRT
                     LVPDAEDPKARAAADAFLTDTLERGHEGVVVKDLAAAYSAGRRGASWLKVKPVHTLDL
                     VVLAVEWGSGRRTGKLSNLHLGARRPDGTFAMLGKTFKGLTDALLDWQTQRLGELATD
                     DDGHVVTVRPELVVEIAYDGLQRSTRYPAGVTLRFARVLRYRDDKTAQEADTVETVLS
                     RRR"
     CDS             complement(8432. .9859)
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAX31523.1"
                     /db_xref="GI:60650898"
                     /translation="MNHPVNGAGERRTTQAREGTQTVAPPRILVVGAGFAGVECVRRL
                     ERRLAPGEAQITLVTPFSYQLYLPLLPQVASGVLTPQSVAVSLRRSRRHRTRIVPGGA
                     IGVDTQAKVCVIRKITDEIVNEPYDYLVLAAGSVTRTFDIPGLLDNARGMKTLAEAAY
                     VRDHVIAQLDLADASHDEAERASRLQFVVVGGGYAGTETAACLQRLTTNAVKHYPRLD
                     PRLIKWHLIDIAPKLMPELGDKLGQAALEVLRKRNIEVSLGVSIAEAGPEEVTFTDGR
                     VLPCRTLIWTAGVAASPLVATLGAETVRGRLAVTPQMRLPGADGVFSLGDAAAVPDLA
                     KGDGAVCPPTAQHAMRQGRVLADNLIASLRHEPLKDYVHKDLGLVVDLGGTDAVSKPL
                     GIELRGLPAQAVARGYHWSALRTNVAKTRVMTNWLLNAVAGDDFVRTGFQSRKPATLR
                     DFEYTDVYLTPEQIKEHTAATVIKH"
```

```
     CDS             complement(9920. .10951)
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAX31524.1"
                     /db_xref="GI:60650899"
                     /translation="MTGRDLTWTDTTSTVDRGRFPDAVTPWEDPAWRAEALAWVTEGL
                     AAHGLTETGPRAVRLRPWSVLVRLAVAGPAPVWFKAVPPAAAFEAGLTEALARWVPAR
                     VLAPLAVEAERGWILVPDGGPVLSEVLDGRPGAPDPGYWEEPLRQYAAMQRELTPYAE
                     AIEALGVPAARPRDLPALFDRLVAGNAALPREDRVALEVLRPRVADWCEELASSGVAD
                     SLDHADLHEKQLFAPVSGRYAFFDWGDALVGHPFCSLLVPARAARERCGPEVLPRLRD
                     AYLEPWTGGGVTAAGLRRAVSLAWRLAALGRAASWGRMFPVPPGGPGVAGDAEGAHWL
                     RELAAAPPL"
     CDS             11239. .12759
                     /codon_start=1
                     /transl_table=11
                     /product="spermine synthase"
                     /protein_id="AAX31525.1"
                     /db_xref="GI:60650900"
                     /translation="MDDAPPRDQGLSSRAAAVLVFGSSAAVLVVEIVALRLLAPYLGL
                     TLETSTLVIGIALTAIALGSWLGGRIADQVDPHRLIAPALGVSGVGVALTPLLLRTTA
                     EWSPALLLLVASATLLVPGALLSAVTPFVTKLRLTSLAETGTVVGRLSGVGTFGAIVG
                     TVLTGFVLVTRLPVSSILIGLGTLLVLGAALVGWQARRWRRATAVALATVVAGTLATG
                     FAPGGCDAETRYHCARVVADPDRDSGRTLVLDGLRHSYVDVEDPEYLKFAYVRAFASV
                     VDTAFPEGEPLTAHHIGGGGLTFPRYLAATRPGTRSLVSEIDPGVVRIDRDRLGLGTP
                     AATGIDVRVEDGRLGLRRLDAGSHDLVVGDAFGGVSAPWHLTTSQALKDVRRALDADG
                     LYVTNLIDHGRLAFARAEVATLAATFPHVALLGQPADIGLDPTASSIGGNMVVVASAR
                     PLDAPAIQEAMDARDVGWRIATGDTLTTWTGNARVLTDDHAPVDQLLQPHPVPAAR"
     CDS             complement(12869. .14650)
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC transporter"
                     /protein_id="AAX31526.1"
                     /db_xref="GI:60650901"
                     /translation="MIGVAPPAYDPAAPESATTLPVGTPTTVRSYVRSLLRRHRRAFT
                     VLIAVNAVAVVASITGPYLLGGLVEDLSAGVTDLHLERTAAIFAVALVVQVLFTRSMR
                     LRGAMLGEEMLADLREDFLVRSVGLPPGVLERAGTGDLLSRITTDIDRLANAMREAVP
                     QLAIGVVWAGLLLGALTVTAPPLALAVLIALPVLIVGCRWYFRRAPSAYRSEAAGYAA
                     VAAMLAETVDAGRTVEAHRLGGRRVALSDRRISQWTAWERYTLFLRSVLFPVINATYV
                     TILGAVLLLGGWFVLEGWLTVGQLTTGALLAQMMVDPIGLILRWYDELQVAQVSLARL
                     VGVRDIEPDAGDAEVGPEGRDVRADEVRFGYREGVDVLHKVSLDVAPGTRLALVGPSG
                     AGKSTLGRLLAGIYAPRTGEVTLGGAELSRMTAERVREHVALVNQEHHVFVGSLRDNL
                     RLAREGAKDAELWASLAAVDADGWAKALEKGLDTEVGSGGFTLTPAQAQQIALARLVL
                     ADPHTLVLDEATSLLDPRAARHLERSLARVLEGRTVVAIAHRLHTAHDADVIAVVEDG
                     RISELGSHDELVAAGGAYASLWRSWHG"
     CDS             complement(14647. .16515)
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC transporter"
                     /protein_id="AAX31527.1"
                     /db_xref="GI:60650902"
                     /translation="MQIRDLPYSDPGDPDVRSGPRFLFWLGRNQLGGQLKSLSWGLLH
                     QLGIAGLPVTVGLAVQAVIDRSGGRLALAGGLIVALGVLIAVGDTMLHRTAVTNWITA
                     AARVQQLLARKTAELGSALTRRVAAGEVVAVSTGDVEKIGWFVEALSRFAAAATALVV
                     ICVGLAVYLPSLGLVVALAMPVLALAVLPLLPRATRRADEQREKAGKATELASDTVAG
                     LRVLRGIGGEELFLGRYRRASQEVRRAAVRSARMWALISAVQVLLPGILLICLVWYGA
                     TLARDGRIDVGQLVTVYSAATLMLFPLRHFEEIAMAYSFSRPSAQRAVRVLSLHRSAQ
                     EATVEGVTPTGDLYDPATGLMAPRGQFTAVVCGDPDEAGRLAERLGGHAETGEEDDKA

  Query Match             10.8%;  Score 107.8;  DB 1;  Length 127559;
  Best Local Similarity   47.6%;  Pred. No. 0.23;
  Matches  445;  Conservative    0;  Mismatches  482;  Indels    8;  Gaps    4;

Qy         42 GACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCC 101
              ||||    ||| ||| ||  ||||  |   |||| ||  |     || ||     | ||| |
Db      96959 GACGGGCTCGGCGACCAGCGGCTCGTCGCGTACGTCACCGGCGGCC-CCGGCACACCGTC 97017

Qy        102 TTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACC 161
                |   ||  | | ||  |  |    ||   ||   |  |||       |  || |
Db      97018 GGCCGCCGCGCTGCGCGCCCACCTGGCGTCCCGGCTGCCCCGGCACATGGTGCCCGGTGA 97077

Qy        162 CATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCA 221
```

```
              | |||   |      |  |   | || |||   ||  | |      |     |||||||
Db      97078 CGTCCTCACCCTGGACGCCCTGCCGCTCACCGCCAACGGCAAGGTGGACCGCACGGCGCT 97137

Qy        222 GCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCG 281
              ||||||| || |  | | |  | || | ||  | |      |  |  ||||| | |   ||
Db      97138 GCCCGGCCCCGGCACCCAGACCGCCGCCCCCGGGCGCGCACCCCAGTCGCCGCAGGAACG 97197

Qy        282 GCGGC--CTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGG 339
              |  ||    |  |     || || | | |||||  |    | || ||     |  ||
Db      97198 GGTGCTGTGCGCCTTGTTCGCCGACGTGCTCGGCCGGGAGACCGTCGGCGTGGACGAGGG 97257

Qy        340 CCTCCGCAGCCCCGGCGG--CCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
               ||  |  ||  |||||   |||   |  | || |   |  |||||||   |   | |
Db      97258 GTTCTTCGACCTGGGCGGTCACTCGCTGCTCGCCACTCGCCTCGCGGCCCGGGTCCGCGC 97317

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
              ||||  |  |   |||    | | ||     | |  || || |    | |   |  |
Db      97318 GGCGCTGGGCGTGGAGATCTCCGTGCGCACCCTGTTCGAGGCGCCGACCCCTGCCCTGCT 97377

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
              |          | ||  |  ||   |||  |  | ||||   ||||   ||  |
Db      97378 CGCGTCGGCGTGCACGGCGGACGCCGCGGCGTACGACCCGTTCGAGACGGTGCTGCCGCT 97437

Qy        518 CCCGCGCGGCTCACGCGGGAGCTC---GTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCA 574
              | ||||       || ||    |  ||||| || |||| ||      || ||
Db      97438 GCGGCGCACGGGCAGCCGGCCACCGCTGTTCTGCGTCCACGCCGGAATGGGCCTGAGCTG 97497

Qy        575 GCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGC 634
              | ||  |   |||    |   || | | ||||    ||| | | ||   | |||||
Db      97498 GGCGTACGCCGGCCTGCTCAGCCATCTGGACGCGGACGTGCCGGTTTACGGACTGCAGGC 97557

Qy        635 CCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTA 694
              |||| |  ||  || || || |  ||    |  |||| ||||||||||  || |   | ||
Db      97558 CCGGAGGCTCACCGCGCCCGGCGGGCTGCCCGGGAGCGTCGAGGAGATGGCTGAGGACTA 97617

Qy        695 CGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTC 754
              || ||      | ||||  |   | |||   ||||| |||    ||   ||   |||||
Db      97618 CGCCGGTGAGATCCGGCGCCTGTGCCCGGATGGGCCGTACCGGCTGCTCGGCTGGTCCTT 97677

Qy        755 CGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTC 814
              ||||||||   |||||  ||   | ||    |  || |    ||| ||  |   ||||
Db      97678 CGGCGGCACGGTCGCCCACGCCGTCGCGACCCGCCTGCAACAGCAGGGCCACACCGTCGA 97737

Qy        815 CCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGG 874
              |  || | | |||||| |||| ||||     || || ||   |  | |    |  ||
Db      97738 ACTCCTCGCCGTCCTCGACGCCTACCCCGTCACCGGGGCCCGGCCCGACGCCGAGGTGGA 97797

Qy        875 CGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCG 934
              |||       |   ||        |||   | | |   |||  | ||  |  ||| ||
Db      97798 CGAACAGCGCATCGTCGCCGACTACCTCGCCCAGCTCGGTTCCCCCGTCGCCCCCGAGCG 97857

Qy        935 CGACGGAGACACGGCGATCAAGCCCGACGACCTCG 969
              | ||  | |   |||    || ||   |||||
Db      97858 CCTCGAGGGCGACGCGTGGCTCCCGGAGTTCCTCG 97892
```

```
RESULT 11
CQ800553/c
LOCUS       CQ800553               71138 bp    DNA     linear   PAT 05-MAY-2004
DEFINITION  Sequence 1 from Patent EP1413626.
ACCESSION   CQ800553
VERSION     CQ800553.1  GI:47057313
KEYWORDS    .
SOURCE      Nonomuraea
  ORGANISM  Nonomuraea
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptosporangineae; Streptosporangiaceae.
REFERENCE   1
  AUTHORS   Donadio,S., Sosio,M. and Beltrametti,F.
  TITLE     Genes and proteins for the biosynthesis of the glycopeptide
            antibiotic A40926
  JOURNAL   Patent: EP 1413626-A 1 28-APR-2004;
            Vicuron Pharmaceuticals, Inc. (US)
```

```
FEATURES             Location/Qualifiers
     source          1. .71138
                     /organism="Nonomuraea"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:83681"
     misc_feature    40. .1140
                     /note="ORF1; negative strandedness"
     misc_feature    1259. .2329
                     /note="ORF2; negative strandedness"
     misc_feature    2558. .5161
                     /note="ORF3; negative strandedness"
     misc_feature    5266. .6231
                     /note="ORF4; negative strandedness"
     misc_feature    7183. .8292
                     /note="ORF5; positive strandedness"
     misc_feature    8320. .8973
                     /note="ORF6; positive strandedness"
     misc_feature    9069. .9659
                     /note="ORF7; positive strandedness"
     misc_feature    9708. .10667
                     /note="ORF8; negative strandedness"
     misc_feature    10670. .11896
                     /note="ORF9; negative strandedness"
     misc_feature    11950. .13419
                     /note="ORF10; negative strandedness"
     misc_feature    13479. .14741
                     /note="ORF11; negative strandedness"
     misc_feature    14823. .16019
                     /note="ORF12; negative strandedness"
     misc_feature    16009. .17163
                     /note="ORF13; negative strandedness"
     misc_feature    17185. .18366
                     /note="ORF14; negative strandedness"
     misc_feature    18462. .18671
                     /note="ORF15; negative strandedness"
     misc_feature    18668. .24259
                     /note="ORF16; negative strandedness"
     misc_feature    24278. .36529
                     /note="ORF17; negative strandedness"
     misc_feature    36760. .39021
                     /note="ORF18; negative strandedness"
     misc_feature    39153. .39851
                     /note="ORF19; negative strandedness"
     misc_feature    40125. .41732
                     /note="ORF20; negative strandedness"
     misc_feature    41772. .42584
                     /note="ORF21; negative strandedness"
     misc_feature    42868. .44130
                     /note="ORF22; negative strandedness"
     misc_feature    44226. .46355
                     /note="ORF23; negative strandedness"
     misc_feature    46632. .48578
                     /note="ORF24; positive strandedness"
     misc_feature    48575. .54868
                     /note="ORF25; positive strandedness"
     misc_feature    54865. .58056
                     /note="ORF26; positive strandedness"
     misc_feature    58152. .58985
                     /note="ORF27; positive strandedness"
     misc_feature    59046. .60641
                     /note="ORF28; positive strandedness"
     misc_feature    60874. .62445
                     /note="ORF29; negative strandedness"
     misc_feature    62887. .63312
                     /note="ORF30; positive strandedness"
     misc_feature    63469. .64587
                     /note="ORF31; positive strandedness"
     misc_feature    64599. .65240
                     /note="ORF32; positive strandedness"
     misc_feature    65237. .66541
                     /note="ORF33; positive strandedness"
     misc_feature    66538. .67335
                     /note="ORF34; positive strandedness"
     misc_feature    67332. .68618
                     /note="ORF35; positive strandedness"
     misc_feature    68685. .69423
                     /note="ORF36; negative strandedness"
     misc_feature    69608. .70894
                     /note="ORF37; positive strandedness"
     misc_feature    71065. .71138
                     /note="attL site, remnant"
ORIGIN

  Query Match            10.8%;  Score 107.6;  DB 6;  Length 71138;
  Best Local Similarity  48.8%;  Pred. No. 0.3;
  Matches  327;  Conservative    0;  Mismatches  334;  Indels    9;  Gaps    1;

Qy        218 CGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGC 277
              ||  |||  || ||||   |||||   ||| |||| |  |   |||     |||    |
Db      19718 CGAGGCCGTGCCCCCGGTGGCCGCCGACCTGGCCGAGGGCGGTCCGCTCGGCGCGCTCGT 19659

Qy        278 CCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGC 337
              | |  | |         ||| || | | |||  | |  |      |  | | | | ||
Db      19658 GCTCGACGACCGGCTGCGCCCCGCCGTCGCCGTCGGCGAGCTCTACCTCACCGGCGCGGC 19599

Qy        338 GGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
              ||    ||| |  |||| | |    |      ||    ||| || |||| | |     |
Db      19598 GGTGGACGCGGAGCCGGGCGACCGGACCCTGGCCTGCCCCTTCGGGGCCACGGGGCGGCG 19539

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
              |  || |  |     |  || |      |   || |||   |         ||  |
Db      19538 CATGCTGCCCACGGGCCTGCTCGCCCGCTGGACAGCCGGGGGCACCCTGGTCGTCGTCGG 19479

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
              || |    |  ||     | |  | | || | | | |||  ||| ||  |
Db      19478 CGAGCGGCGGGGATCGAGCGGCTCGGTGAAGACCGGCACGGGCGACTTCGAGGTGCTGCT 19419

Qy        518 CCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCC 577
              ||||  |  |  |||| | |      | || || |||  | || |  ||    ||
Db      19418 GCCGCTGCGAGCCGGCGGTAACCGTCCGCCGCTGTACTGCGTCCATGCGAGCGGAGGGCT 19359

Qy        578 GCTCTATCCCCTGTCTGCCGCGCTCGG---------CACGGATTACCCGTTCTTCGCCTT 628
              |  ||    |   |  ||  |||  |         | |  |  | ||| ||| || | |
Db      19358 GAGCTGGAACTACGCGCCCCTGCTGCGGAGCCTGCCCCCCAACCAGCCGGTCTACGGCGT 19299

Qy        629 CCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGC 688
              ||||| || ||| | |    |      ||      |  || |||||||||| |||||
Db      19298 GCAGGCGCGCGGCCTGGCTCGTACGGAGCCGCTGGCGGCCGGCGTCGAGGAGATGGCGGC 19239

Qy        689 CCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCT 748
              | |||||||||     | ||       ||||  |||| |||  | |    |||
Db      19238 CGATTACGTCGAGCAGATCCGCGCCGTGCAGCCGACCGGGCCGTACCACCTCCTCGGGTG 19179

Qy        749 GTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCG 808
              |||    |||||    |||||    |||||||||  ||  |||      ||||||||||
Db      19178 GTCACTGGGCGGGCGGATCGCGCAGGAGATGGCCAGGGTGCTGGAGGAGCAAGGGGAGCA 19119

Qy        809 CGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGAC 868
              |||  || ||| | | |||||||| |||||||| | | || ||||      | |    |
Db      19118 GGTCGGCCTGCTCGCCCTGCTCGACGCCTATCCCACCGACGTGGGCAGGCTCCGGCGCCC 19059

Qy        869 GCCGGGCGAC 878
              ||  ||||||
Db      19058 GCGCGGCGAC 19049


RESULT 12
CQ813424/c
LOCUS       CQ813424               71138 bp    DNA     linear   PAT 24-MAY-2004
DEFINITION  Sequence 1 from Patent WO2004038025.
ACCESSION   CQ813424
VERSION     CQ813424.1  GI:47602716
KEYWORDS    .
SOURCE      Nonomuraea
  ORGANISM  Nonomuraea
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
```

```
          Streptosporangineae; Streptosporangiaceae.
REFERENCE 1
  AUTHORS Donadio,S., Sosio,M. and Beltrametti,F.
  TITLE   Genes and proteins for the biosynthesis of the glycopeptide
          antibiotic a40926
  JOURNAL Patent: WO 2004038025-A 1 06-MAY-2004;
          Vicuron Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .71138
                     /organism="Nonomuraea"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:83681"
     misc_feature    40. .1140
                     /note="ORF1; negative strandedness"
     misc_feature    1259. .2329
                     /note="ORF2; negative strandedness"
     misc_feature    2558. .5161
                     /note="ORF3; negative strandedness"
     misc_feature    5266. .6231
                     /note="ORF4; negative strandedness"
     misc_feature    7183. .8292
                     /note="ORF5; positive strandedness"
     misc_feature    8320. .8973
                     /note="ORF6; positive strandedness"
     misc_feature    9069. .9659
                     /note="ORF7; positive strandedness"
     misc_feature    9708. .10667
                     /note="ORF8; negative strandedness"
     misc_feature    10670. .11896
                     /note="ORF9; negative strandedness"
     misc_feature    11950. .13419
                     /note="ORF10; negative strandedness"
     misc_feature    13479. .14741
                     /note="ORF11; negative strandedness"
     misc_feature    14823. .16019
                     /note="ORF12; negative strandedness"
     misc_feature    16009. .17163
                     /note="ORF13; negative strandedness"
     misc_feature    17185. .18366
                     /note="ORF14; negative strandedness"
     misc_feature    18462. .18671
                     /note="ORF15; negative strandedness"
     misc_feature    18668. .24259
                     /note="ORF16; negative strandedness"
     misc_feature    24278. .36529
                     /note="ORF17; negative strandedness"
     misc_feature    36760. .39021
                     /note="ORF18; negative strandedness"
     misc_feature    39153. .39851
                     /note="ORF19; negative strandedness"
     misc_feature    40125. .41732
                     /note="ORF20; negative strandedness"
     misc_feature    41772. .42584
                     /note="ORF21; negative strandedness"
     misc_feature    42868. .44130
                     /note="ORF22; negative strandedness"
     misc_feature    44226. .46355
                     /note="ORF23; negative strandedness"
     misc_feature    46632. .48578
                     /note="ORF24; positive strandedness"
     misc_feature    48575. .54868
                     /note="ORF25; positive strandedness"
     misc_feature    54865. .58056
                     /note="ORF26; positive strandedness"
     misc_feature    58152. .58985
                     /note="ORF27; positive strandedness"
     misc_feature    59046. .60641
                     /note="ORF28; positive strandedness"
     misc_feature    60874. .62445
                     /note="ORF29; negative strandedness"
     misc_feature    62887. .63312
                     /note="ORF30; positive strandedness"
     misc_feature    63469. .64587
                     /note="ORF31; positive strandedness"
     misc_feature    64599. .65240
                     /note="ORF32; positive strandedness"
     misc_feature    65237. .66541
                     /note="ORF33; positive strandedness"
     misc_feature    66538. .67335
                     /note="ORF34; positive strandedness"
     misc_feature    67332. .68618
                     /note="ORF35; positive strandedness"
     misc_feature    68685. .69423
                     /note="ORF36; negative strandedness"
     misc_feature    69608. .70894
                     /note="ORF37; positive strandedness"
     misc_feature    71065. .71138
                     /note="attL site, remnant"
ORIGIN

  Query Match             10.8%;  Score 107.6;  DB 6;  Length 71138;
  Best Local Similarity   48.8%;  Pred. No. 0.3;
  Matches  327;  Conservative    0;  Mismatches  334;  Indels    9;  Gaps    1;

Qy        218 CGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGC 277
              ||  |||  || ||||  |||||  ||| |||| | |  |||  |||  |
Db      19718 CGAGGCCGTGCCCCCGGTGGCCGCCGACCTGGCCGAGGGCGGTCCGCTCGGCGCGCTCGT 19659

Qy        278 CCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGC 337
              | | | |          ||| || | | |||  | | |      |  | | | | | ||
Db      19658 GCTCGACGACCGGCTGCGCCCCGCCGTCGCCGTCGGCGAGCTCTACCTCACCGGCGCGGC 19599

Qy        338 GGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
              ||    ||| |  |||| | | |    |      ||    ||| || |||| | |     |
Db      19598 GGTGGACGCGGAGCCGGGCGACCGGACCCTGGCCTGCCCCTTCGGGGCCACGGGGCGGCG 19539

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
              |  || |  |    |  || |    |  || |||  |      ||  |  |
Db      19538 CATGCTGCCCACGGGCCTGCTCGCCCGCTGGACAGCCGGGGGCACCCTGGTCGTCGTCGG 19479

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
              ||  |    |  ||     |  |  |  | ||  | | | |||  |||  ||  |
Db      19478 CGAGCGGCGGGGATCGAGCGGCTCGGTGAAGACCGGCACGGGCGACTTCGAGGTGCTGCT 19419

Qy        518 CCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCC 577
              ||||   |  |  |||| | |        | || || |||  | || |  ||    ||
Db      19418 GCCGCTGCGAGCCGGCGGTAACCGTCCGCCGCTGTACTGCGTCCATGCGAGCGGAGGGCT 19359

Qy        578 GCTCTATCCCCTGTCTGCCGCGCTCGG---------CACGGATTACCCGTTCTTCGCCTT 628
              | ||    |   |  ||  |||  |          | |  |  | ||| ||| | |
Db      19358 GAGCTGGAACTACGCGCCCCTGCTGCGGAGCCTGCCCCCCAACCAGCCGGTCTACGGCGT 19299

Qy        629 CCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGC 688
              ||||| || ||| | |   |    |    ||       |  || ||||||||||| |||||
Db      19298 GCAGGCGCGCGGCCTGGCTCGTACGGAGCCGCTGGCGGCCGGCGTCGAGGAGATGGCGGC 19239

Qy        689 CCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCT 748
              | |||||||||       | ||           ||||  ||||| |||  | |     |||
Db      19238 CGATTACGTCGAGCAGATCCGCGCCGTGCAGCCGACCGGGCCGTACCACCTCCTCGGGTG 19179

Qy        749 GTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCG 808
              |||    |||||    |||||    ||||||||| ||  |||      ||||||||||
Db      19178 GTCACTGGGCGGGCGGATCGCGCAGGAGATGGCCAGGGTGCTGGAGGAGCAAGGGGAGCA 19119

Qy        809 CGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGAC 868
              |||  || ||| | | |||||||| |||||||| | | || ||||       | |     |
Db      19118 GGTCGGCCTGCTCGCCCTGCTCGACGCCTATCCCACCGACGTGGGCAGGCTCCGGCGCCC 19059

Qy        869 GCCGGGCGAC 878
              ||  ||||||
Db      19058 GCGCGGCGAC 19049


RESULT 13
NSP561198/c
LOCUS       NSP561198              89153 bp    DNA     linear   BCT 15-APR-2005
```

```
DEFINITION  Actinomadura sp. ATCC 39727 gene cluster for biosynthesis of
            glycopeptide antibiotic A40926, strain ATCC 39727.
ACCESSION   AJ561198
VERSION     AJ561198.1  GI:32487222
KEYWORDS    ABC transporter; acyltransferase; beta-hydroxylase; dalbavancin;
            dbv1; dbv10; dbv11; dbv12; dbv13; dbv14; dbv15; dbv16; dbv17;
            dbv18; dbv19; dbv2; dbv20; dbv21; dbv22; dbv23; dbv24; dbv25;
            dbv26; dbv27; dbv28; dbv29; dbv3; dbv30; dbv31; dbv32; dbv33;
            dbv34; dbv35; dbv36; dbv37; dbv4; dbv5; dbv6; dbv7; dbv8; dbv9;
            DpgA protein; DpgB protein; DpgC protein; DpgD protein;
            glycosyltransferase; halogenase; hexose oxidase; HpgT
            aminotransferase; hydroxymandelate oxidase; hydroxymandelate
            synthase; integral membrane ion antiporter; m ethyltransferase;
            mannosyltransferase; monooxygenase; non-ribosomal peptide
            synthetase; ORF1; ORF10; ORF11; ORF12; ORF13; ORF14; ORF15; ORF16;
            ORF17; ORF18; ORF2; ORF3; ORF4; ORF5; ORF6; ORF7; ORF8; ORF9; OxyA
            monooxygenase; OxyB monooxigenase; oxygenase OxyC; prephenate
            dehydrogenase; regulator LuxR family; regulator StrR family;
            response regulator; sensory kinase; type II thioesterase; VanY-type
            carboxypeptidase.
SOURCE      Nonomuraea sp. ATCC 39727
  ORGANISM  Nonomuraea sp. ATCC 39727
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptosporangineae; Streptosporangiaceae; Nonomuraea.
REFERENCE   1
  AUTHORS   Sosio,M., Stinchi,S., Beltrametti,F., Lazzarini,A. and Donadio,S.
  TITLE     The gene cluster for the biosynthesis of the glycopeptide
            antibiotic A40926 by nonomuraea species
  JOURNAL   Chem. Biol. 10 (6), 541-549 (2003)
   PUBMED   12837387
REFERENCE   2  (bases 1 to 89153)
  AUTHORS   Donadio,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2003) Donadio S., Microbial Technologies, Vicuron
            Pharmaceuticals, via R. Lepetit 34, 21040 Gerenzano, ITALY
FEATURES             Location/Qualifiers
     source          1. .89153
                     /organism="Nonomuraea sp. ATCC 39727"
                     /mol_type="genomic DNA"
                     /strain="ATCC 39727"
                     /db_xref="taxon:93944"
                     /note="[Actinomadura] sp. ATCC 39727 = cf. Nonomuraea sp."
     CDS             complement(<1. .141)
                     /function="putative Ser/Thr protein kinase"
                     /note="ORF1
                     outside dalbavancin cluster"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD91190.1"
                     /db_xref="GI:32487223"
                     /db_xref="GOA:Q7WZ96"
                     /db_xref="InterPro:IPR000719"
                     /db_xref="UniProt/TrEMBL:Q7WZ96"
                     /translation="MGGSAELLPGDPARLGDYWLAGRLGSGGQGVVYEAYDEGGRRVA
                     IKI"
     CDS             complement(347. .1207)
                     /function="putative hydrolase"
                     /note="ORF2
                     outside dalbavancin cluster"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD91191.1"
                     /db_xref="GI:32487224"
                     /db_xref="GOA:Q7WZ95"
                     /db_xref="InterPro:IPR000073"
                     /db_xref="InterPro:IPR000379"
                     /db_xref="UniProt/TrEMBL:Q7WZ95"
                     /translation="MTRHTTVQIQGLPVFYREAGDPANPTLVLLHGFPSSSAMFRNLI
                     ADLADSYHLIAPDHVGFGQSAAPSVEEFPYSFDKLTEITLELLDTLNIDRFALYIHDY
                     GAPIGLRIASRHPERVSAIITQSGNAYLDGFTPFWEVLFAHAKDRATHEPAVRELLTA
                     DATRWQYTHGVPADLLERVSPDSWVLDQAYLDRPGNKEVQLQLFWDYQFNLDGYPAFQ
                     EYFRTHRPPVLVAWGANDEIFGAAGAEAYARDLPDAEIHLLDAGHFALETHGREIAGL
                     IRDFLRRRLP"
     CDS             complement(1797. .2627)
                     /function="putative DNA binding protein"
                     /note="ORF3
                     outside dalbavancin cluster"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD91192.1"
                     /db_xref="GI:32487225"
                     /db_xref="GOA:Q7WZ94"
                     /db_xref="InterPro:IPR001387"
                     /db_xref="UniProt/TrEMBL:Q7WZ94"
                     /translation="MSRQDELREFLRSRRSRVAPRDVGLPARAGGRRVRGLRREELAL
                     AAGVSADYYTRMEQGRAGNVSDQVLEAVAGVLRLTAEEQRHLRELIGDRHHARPRPPA
                     RPRAALRMMLDMLGPAPALLMTGVLDILALNRMGAALLTDFEAMPRPERNEIRWIFLD
                     PRARDVFPDWEEVAVEAAAWLRGALGRSGDDPRPQQLVDELSARSADFARHWADHRVT
                     YCTYGTKRLRHPEVGVMTLNWESFAPTADPDLYLVVYTAPTGSPSEERLAALAGGQVS
                     "
     CDS             complement(3225. .4181)
                     /function="putative regulator"
                     /note="ORF4
                     outside dalbavancin cluster"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD91193.1"
                     /db_xref="GI:32487226"
                     /db_xref="InterPro:IPR003770"
                     /db_xref="UniProt/TrEMBL:Q7WZ93"
                     /translation="MNIEDLLRETLSEMADEERPPAPSRFLQGRSLQAREHRPRRRGL
                     ALMAAAGVVALAVGATFVVRGLSPGTSGEAAGQPDAVASVVPEQRPAITVIVPPGMRL
                     AQLLKRLSSVTGRPVAEFERAARDGAALGLPPYAKGRLEGFAAPGSYEISPAMSPGEI
                     LGAMVTGFGRLAEETGLVAGAGRAGRTPLEILTIASIVQAESFRPEDMPKIARVLHNR
                     LDRKMRLQVDSTVLYGLGKYGSAATPEDVRSPSPYNTYRRPGLPPGPIGSPGADAVRA
                     ALEPAAGRWLYYVATDPRTGAMKFATTQTEYAALVEERDRNQ"
     CDS             complement(4178. .4690)
                     /note="ORF5
                     outside dalbavancin cluster"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD91194.1"
                     /db_xref="GI:32487227"
                     /db_xref="GOA:Q7WZ92"
                     /db_xref="InterPro:IPR007627"
                     /db_xref="InterPro:IPR007630"
                     /db_xref="UniProt/TrEMBL:Q7WZ92"
                     /translation="MDPGFEAAFRDFVVGRSGALFRTAYLLTGERHAAEDLVQSALAK
                     TAAKWRGLRDPAAIEGYVRRVMYHEQVSWWRRRSRVAEVSTGWLPDQARDGHADGADL
                     RVMMRAALGRLTPRQRTILVLRYFEDLSETEIARVLGIRVGSVRSQIHRSLERLKQTA
                     PELTIMREFG"
     CDS             4838. .6202
                     /note="ORF6
                     outside dalbavancin cluster"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD91195.1"
                     /db_xref="GI:32487228"
                     /db_xref="GOA:Q7WZ91"
                     /db_xref="InterPro:IPR004347"
                     /db_xref="InterPro:IPR004989"
                     /db_xref="InterPro:IPR011389"
                     /db_xref="UniProt/TrEMBL:Q7WZ91"
                     /translation="MAMDRRIFGLENEYGITCTFKGQRRLSPDEVARYLFRKVVSWGR
                     SSNVFLGNGARLYLDVGSHPEYATPECDNVVELVTHDKAGERIVEGLAVDAERRLREE
                     GIAGDIYLFKNNTDSAGNSYGCHENYCVGRHGEFGRLADVLIPFLVTRQVVCGAGKVL
                     QTPRGAVYCLSQRAEHIWEGVSSATTRSRPIINTRDEPHGDAERFRRLHVIVGDSNMS
                     ETTTLLKIGATDLVLRMIEAGVVMRDLSLENPIRAIREVSHDMTGRRKVRLANGREAS
                     ALEIQQEYLSKAADFVDRRGGDATALRVLELWERTLRAVETGDLDLVSREIDWVTKYQ
                     MIERYRSKYDLPLSSSRVAQLDLAYHDVHRRRGLFYLLQRKGLVERVASDVKIFEAKS
```

```
                     VPPQTTRARLRGEFIRKAQEKRRDFTVDWVHLKLNDQAQRTVLCKDPFRSVDERVDRL
                     IAGM"
     gene            complement(6242. .7342)
                     /gene="dbv1"
     CDS             complement(6242. .7342)
                     /gene="dbv1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative hydroxymandelate oxidase"
                     /protein_id="CAD91196.1"
                     /db_xref="GI:32487229"
                     /db_xref="GOA:Q7WZ90"
                     /db_xref="InterPro:IPR003009"
                     /db_xref="InterPro:IPR008259"
                     /db_xref="UniProt/TrEMBL:Q7WZ90"
                     /translation="MHESPVCLAEYEEIAAKVLPADVRDFIDGGSGREQTLRANRAAF
                     DRVFLVPRVLQDVSACSTRATLLGHPATMPVAVAPVAYHRLVHPDGELATARAARDAG
                     VPFTVSTLSSVPVEDVTALGGHVWFQLYCLREHAATLGLIRRAEDAGCRALMLTLDVP
                     WMGRRPRDIRNRFRLPPHVRPVHLTANSGTEAHRGASGGSALAAHTAMELSAAVDWSY
                     LETLRAASGLPLVVKGILHPEDARRAADLGIDGIVVSNHGGRQLDGAVASLDALPGVA
                     ESVGGRCEIMLDGGVRSGADVLKALALGASGVLVGRPVIWGLAADGERGVRTVLGLLG
                     AEIEDGLGLAGCGDVAAAQALRTTRPGAGFVS"
     gene            complement(7461. .8531)
                     /gene="dbv2"
     CDS             complement(7461. .8531)
                     /gene="dbv2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative hydroxymandelate synthase"
                     /protein_id="CAD91197.1"
                     /db_xref="GI:32487230"
                     /db_xref="GOA:Q7WZ89"
                     /db_xref="InterPro:IPR004360"
                     /db_xref="InterPro:IPR005956"
                     /db_xref="UniProt/TrEMBL:Q7WZ89"
                     /translation="MESLPPLAVDYVEMYVADLKVATLPWTDEYRFAVVGTANASDHR
                     SVALRQGRITLVLTQATSDGHPASAYVRTHGDGVADIALRTPDVDVVFTHAVAAGARP
                     VRSPSRHPGPGPACSAAIGGFGDVVHTLVQRDPGDDPGLPVGFSEAPSAAESGADAAE
                     LLDIDHFAVCLPTGDLDIITDFYVATLGFSETFKERIEVGTQAMESKVVQSASGAVTL
                     TLIEPDPMAEAGQIDMFLERHAGAGVQHVAFSSSDAVHAVNTLSERGVRFLSTPGSYY
                     DLLESRIQIRGHTVDQLRATGLLADEDHGGQLFQIFTASTHPRETLFFEVIERQGART
                     FGGANIKALYEAVEVARSQQRA"
     gene            complement(8760. .11363)
                     /gene="dbv3"
     CDS             complement(8760. .11363)
                     /gene="dbv3"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulator LuxR family"
                     /protein_id="CAD91198.1"
                     /db_xref="GI:32487231"
                     /db_xref="GOA:Q7WZ88"
                     /db_xref="InterPro:IPR000792"

  Query Match             10.8%;  Score 107.6;  DB 1;  Length 89153;
  Best Local Similarity   48.8%;  Pred. No. 0.28;
  Matches  327;  Conservative    0;  Mismatches  334;  Indels    9;  Gaps    1;

Qy        218 CGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGC 277
              ||  |||  || ||||   |||||   ||| |||| |  |   |||     |||    |
Db      25920 CGAGGCCGTGCCCCCGGTGGCCGCCGACCTGGCCGAGGGCGGTCCGCTCGGCGCGCTCGT 25861

Qy        278 CCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGC 337
                | |  | |         ||| || | | |||  | |  |      |  | | | | | ||

Db      25860 GCTCGACGACCGGCTGCGCCCCGCCGTCGCCGTCGGCGAGCTCTACCTCACCGGCGCGGC 25801

Qy        338 GGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
              ||    ||| |  ||||  | |    |     ||    ||| || |||| | |   |

Db      25800 GGTGGACGCGGAGCCGGGCGACCGGACCCTGGCCTGCCCCTTCGGGGCCACGGGGCGGCG 25741

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
                | || |   |      |  || |      |   || |||   |        ||  |

Db      25740 CATGCTGCCCACGGGCCTGCTCGCCCGCTGGACAGCCGGGGGCACCCTGGTCGTCGTCGG 25681
```

```
Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
              ||  |    |   ||     |  |  |  | ||  | | | |||  |||  ||  |

Db      25680 CGAGCGGCGGGGATCGAGCGGCTCGGTGAAGACCGGCACGGGCGACTTCGAGGTGCTGCT 25621

Qy        518 CCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCC 577
              ||||   |  |  |||| | |         | || || |||  | || |  ||    ||

Db      25620 GCCGCTGCGAGCCGGCGGTAACCGTCCGCCGCTGTACTGCGTCCATGCGAGCGGAGGGCT 25561

Qy        578 GCTCTATCCCCTGTCTGCCGCGCTCGG---------CACGGATTACCCGTTCTTCGCCTT 628
              |  ||     |    |  ||  |||  |         | |  |  | ||| ||| || | |

Db      25560 GAGCTGGAACTACGCGCCCCTGCTGCGGAGCCTGCCCCCCAACCAGCCGGTCTACGGCGT 25501

Qy        629 CCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGC 688
              ||||| || ||| | |   |    |    ||       |  || ||||||||||| |||||

Db      25500 GCAGGCGCGCGGCCTGGCTCGTACGGAGCCGCTGGCGGCCGGCGTCGAGGAGATGGCGGC 25441

Qy        689 CCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCT 748
              | |||||||||       | ||           ||||  ||||| |||  | |    |||

Db      25440 CGATTACGTCGAGCAGATCCGCGCCGTGCAGCCGACCGGGCCGTACCACCTCCTCGGGTG 25381

Qy        749 GTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCG 808
              |||    |||||    |||||    |||||||| ||  |||      ||||||||||

Db      25380 GTCACTGGGCGGGCGGATCGCGCAGGAGATGGCCAGGGTGCTGGAGGAGCAAGGGGAGCA 25321

Qy        809 CGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGAC 868
              |||  || ||| | | |||||||| |||||||| | | || ||||      | |    |

Db      25320 GGTCGGCCTGCTCGCCCTGCTCGACGCCTATCCCACCGACGTGGGCAGGCTCCGGCGCCC 25261

Qy        869 GCCGGGCGAC 878
              ||  ||||||
Db      25260 GCGCGGCGAC 25251
```

```
RESULT 14
AR636660
LOCUS       AR636660               7110 bp    DNA     linear   PAT 20-APR-2005
DEFINITION  Sequence 26 from patent US 6855513.
ACCESSION   AR636660
VERSION     AR636660.1  GI:62769686
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7110)
  AUTHORS   Whiteley,M., Lee,K.M., Greenberg,E.P. and Muh,U.
  TITLE     Quorum sensing signaling in bacteria
  JOURNAL   Patent: US 6855513-A 26 15-FEB-2005;
            University of Iowa Research Foundation and Vertex Pharmaceuticals
            (San Diego) LLC; Iowa City, IA
FEATURES             Location/Qualifiers
     source          1. .7110
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             10.6%;  Score 106.4;  DB 6;  Length 7110;
  Best Local Similarity   50.8%;  Pred. No. 0.97;
  Matches  308;  Conservative    0;  Mismatches  291;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||             ||   ||

Db       6007 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6066

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| |||||  |  |  | |  || | ||  |  |   ||| |||      |||  |

Db       6067 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6126

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
              | |       ||   | | |  ||||     |||    |||  | | |  | |  ||

Db       6127 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6182

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
```

```
Db       6183 GACGCCGACCGTGGAGGGCCTGGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6242

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              |||| ||| ||   |||    |  |||||  |||  |    ||||   || |||
Db       6243 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6302

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
                || | || ||       |   || |||     ||| ||     ||||  | ||
Db       6303 GCTCGGCGGCCACGTGCTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6362

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |   | |||  |  ||  | ||||
Db       6363 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6422

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              ||| |||||  |||  |||| ||||   | | || | |      |||  ||| || ||
Db       6423 GGACATCGCCGCCAGTTACCTCGCGGCCATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6482

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| ||||||||||||||| || | | |
Db       6483 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6542

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |   | ||   ||| |||  ||| || |||||||| ||    |||  | |
Db       6543 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6602

Qy        854 CCTCGC 859
              || |||
Db       6603 CCACGC 6608
```

```
RESULT 15
AX097461
LOCUS       AX097461               7110 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 26 from Patent WO0118248.
ACCESSION   AX097461
VERSION     AX097461.1  GI:13514043
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1
  AUTHORS   Whiteley,M., Lee,K.M., Greenberg,E.P. and Muh,U.
  TITLE     Quorum sensing signaling in bacteria
  JOURNAL   Patent: WO 0118248-A 26 15-MAR-2001;
            University of Iowa Research Foundation Inc. (US) ; Quorum Sciences
            Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .7110
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:287"
ORIGIN

  Query Match             10.6%;  Score 106.4;  DB 6;  Length 7110;
  Best Local Similarity   50.8%;  Pred. No. 0.97;
  Matches  308;  Conservative    0;  Mismatches  291;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||           ||   ||
Db       6007 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6066

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| |||||  |  |  |  |  || | ||  |  |   ||| |||     |||  |
Db       6067 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6126

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
               | |       ||   |  | |   ||||     |||    |||  | | |  | |  ||
Db       6127 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6182

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
```

```
Db       6183 GACGCCGACCGTGGAGGGCCTGGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6242

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              |||| ||| ||   |||    |  |||||  |||  |    ||||   || |||
Db       6243 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6302

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
                || | || ||       |   || |||     ||| ||     ||||  | ||
Db       6303 GCTCGGCGGCCACGTGCTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6362

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |   | |||  |  ||  | ||||
Db       6363 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6422

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              ||| |||||  |||  |||| ||||   | | || | |      |||  ||| || ||
Db       6423 GGACATCGCCGCCAGTTACCTCGCGGCCATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6482

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| ||||||||||||||| || | | |
Db       6483 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6542

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |   | ||   ||| |||  ||| || |||||||| ||    |||  | |
Db       6543 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6602

Qy        854 CCTCGC 859
              || |||
Db       6603 CCACGC 6608
```

```
Search completed: April 28, 2006, 02:39:38
Job time : 5544.08 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 482.212 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_55000_55999
Perfect score: 1000
Sequence: 1 tcacggcgatggtctgcgat.....,...cccccgatgcagctcccgcg 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	76994	12	ADP64454	Adp64454 Sorangium
	2	120.2	12.0	4296	14	ACL66828	Acl66828 M. xanthu
c	3	120.2	12.0	17727	14	ACL64689	Acl64689 M. xanthu
	4	110	11.0	1121	8	ACA25936	Aca25936 Prokaryot
	5	107.8	10.8	90597	10	ADJ72363	Adj72363 Streptomy
	6	107.8	10.8	90600	6	ABQ78872	Abq78872 S. roseos
c	7	107.6	10.8	71138	12	ADO13830	Ado13830 dbv gene
	8	106.4	10.6	7059	13	ADS14647	Ads14647 Pseudomon
	9	106.4	10.6	7110	4	AAF81361	Aaf81361 Quorum se
	10	102.8	10.3	5862	10	AAD54224	Aad54224 Streptomy
	11	102.8	10.3	52101	10	AAD54217	Aad54217 Streptomy
c	12	101.6	10.2	1209	11	ABD02736	Abd02736 Pseudomon
	13	101.6	10.2	1260	11	ABD02659	Abd02659 Pseudomon
	14	101.6	10.2	6858	11	ABD02615	Abd02615 Pseudomon
	15	100.8	10.1	10002	8	ACA26601	Aca26601 Prokaryot
	16	98.6	9.9	24459	14	ACL68541	Acl68541 M. xanthu
c	17	98.6	9.9	32241	14	ACL64784	Acl64784 M. xanthu
	18	98	9.8	1002	14	ACL67076	Acl67076 M. xanthu
	19	98	9.8	1008	14	ACL63654	Acl63654 M. xanthu
	20	98	9.8	6297	10	AAD54236	Aad54236 Streptomy
	21	98	9.8	50543	10	AAD54230	Aad54230 Streptomy
	22	97.2	9.7	2196	14	ACL70073	Acl70073 M. xanthu
	23	97.2	9.7	2200	14	ACL64015	Acl64015 M. xanthu
c	24	97	9.7	462	11	ABD02695	Abd02695 Pseudomon
c	25	96.2	9.6	3024	14	AEB57391	Aeb57391 Silk elas
c	26	96.2	9.6	37856	3	AAA11992	Aaa11992 S. cellul
	27	96	9.6	3957	5	AAA09686	Aaa09686 HSV-2 imm
	28	96	9.6	3957	10	ADG75121	Adg75121 Human her
	29	96	9.6	154746	6	AAD25519	Aad25519 Human her
c	30	96	9.6	154746	6	AAD25519	Aad25519 Human her
	31	95.6	9.6	2937	14	ACL70212	Acl70212 M. xanthu
	32	95.6	9.6	3550	14	ACL64061	Acl64061 M. xanthu
	33	93.2	9.3	1337	2	AAZ17263	Aaz17263 Human gen
	34	92.4	9.2	3978	3	AAC55785	Aac55785 Type I po
	35	92.4	9.2	3978	10	ADE10204	Ade10204 S. lavend
	36	92.4	9.2	12249	3	AAC55840	Aac55840 Complete
	37	92.4	9.2	12249	10	ADE10259	Ade10259 S. lavend
	38	92.4	9.2	18331	3	AAC55857	Aac55857 Complete
	39	92.4	9.2	18331	10	ADE10266	Ade10266 S. lavend
c	40	92.2	9.2	3133	10	ADC86738	Adc86738 Human GPC
c	41	91.6	9.2	12733	6	ABK98631	Abk98631 Vector pE
c	42	91.6	9.2	12733	9	ACD13882	Acd13882 L. lactis
c	43	91.6	9.2	12739	6	ABK98592	Abk98592 Vector pE
c	44	91.6	9.2	12739	9	ACD13843	Acd13843 Plasmid p
	45	91.4	9.1	73882	13	ADS73531	Ads73531 tcp gene

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

```
PS   Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC   The present invention describes an isolated recombinant polynucleotide
CC   (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC   synthase (PKS) protein or a fragment comprising one or more domains of
CC   the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC   host cell (III) comprising (II), or (I) integrated into the cell
CC   chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC   disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC   the native disorazole PKS by the inactivation of one or more domain of
CC   the disorazole PKS and/or addition of one or more domain of a non-
CC   disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC   expression system (VI) capable of producing a disorazole synthase domain
CC   in a host cell, where the system comprises an encoding sequence for a
CC   disorazole polyketide synthase domain, and the encoding sequence is
CC   operably linked to control sequences effective in the cell to produce RNA
CC   that is translated into the domain; (7) host cell modified to contain
CC   (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC   in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC   eliminate production of disorazole; (9) an isolated polypeptide encoded
CC   by (I); (10) hybrid polyketide synthase comprising one or more
CC   polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC   disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC   of at least 200 base pairs with a sequence identical or substantially
CC   identical to a protein encoding region of the 77294 base pair sequence of
CC   SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC   inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC   (I) is useful for producing disorazole polyketides. (III) is useful for
CC   producing a polyketide, which involves growing (III) under conditions,
CC   where a polyketide synthesised by a PKS comprising a protein encoded by
CC   the polynucleotide molecule is produced in the cell. (I) is useful for
CC   producing libraries of PKSs, where the polyketide producing colonies are
CC   identified and isolated, and the produced polyketides are used
CC   collectively in a panel to represent a library or assessed individually
CC   for activity. The disorazole polyketide produced by (I) is useful in
CC   molecular biology, chemistry, recombinant DNA technology, human and
CC   veterinary medicine and agricultural applications. The disorazole
CC   polyketide produced by (I) is formulated as a pharmaceutical composition,
CC   useful for its activity against microorganisms, in a mammal. The present
CC   sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC   disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 1000;  DB 12;  Length 76994;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-130;
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCACGGCGATGGTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54700 TCACGGCGATGGTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGG 54759

Qy         61 CGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54760 CGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGC 54819

Qy        121 AGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54820 AGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGC 54879

Qy        181 TCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54880 TCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCG 54939

Qy        241 CGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54940 CGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGG 54999

Qy        301 CGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55000 CGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCT 55059

Qy        361 CCGCGGCCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55060 CCGCGGCCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGG 55119

Qy        421 CGCCTGATCCGCGCGCGGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55120 CGCCTGATCCGCGCGCGGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCG 55179

Qy        481 AGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55180 AGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCT 55239

Qy        541 CGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55240 CGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGC 55299

Qy        601 TCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55300 TCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGC 55359

Qy        661 CCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55360 CCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTC 55419

Qy        721 CGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55420 CGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGG 55479

Qy        781 CCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55480 CCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATC 55539

Qy        841 CCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55540 CCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCC 55599

Qy        901 TGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55600 TGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCG 55659

Qy        961 ACGACCTCGCCGGGCTCCCCCCCCCCGATGCAGCTCCCGCG 1000
              |||||||||||||||||||||||||||||||||||||||||
Db      55660 ACGACCTCGCCGGGCTCCCCCCCCCCGATGCAGCTCCCGCG 55699

RESULT 2
ACL66828
ID   ACL66828 standard; DNA; 4296 BP.
XX
AC   ACL66828;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus gene sequence, seq id 3291.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression;
KW   gene; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,  Hinkle GJ,  Slater SC,  Wiegand RC;
XX
DR   WPI; 2005-028716/03.
```

XX
PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 2; SEQ ID NO 3291; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC set of about 7842 genes or partial genes from the genome of the bacterium
CC Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from USPTO
XX
SQ Sequence 4296 BP; 581 A; 1497 C; 1606 G; 612 T; 0 U; 0 Other;

Query Match 12.0%; Score 120.2; DB 14; Length 4296;
Best Local Similarity 52.2%; Pred. No. 2.1e-08;
Matches 314; Conservative 0; Mismatches 283; Indels 4; Gaps 2;

```
Qy       326 CGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGC 385
Db      3605 CGCTGGGCGCGGACCTGACGCTGGCCACGCTGTTCGAGGCTCCCACGCTGGAGCAGTGCG 3664

Qy       386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTC-CGGCGCCTGATCCGCGCGCGGGGGACG 444
Db      3665 CCGCGCTGGTGCGTGAGGCCGCCGGGATTCCCTTCACGCCGGACGCGGGTCCGGGTGACG 3724

Qy       445 CACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGC 504
Db      3725 CGGCGCCGGCTGCGCAGGCCCCGGGCGTCAAGGCGGTGCCGAAG---AGCTGGTCTCCGC 3781

Qy       505 TCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTG 564
Db      3782 TGGTCACCATCCAGAAGGGCGGCAATACCCTCCCCTTCTTCTGCGTGCACGGCGCGGGCG 3841

Qy       565 GGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCG 624
Db      3842 GCAACGTCCTCAACTTCCGCGAGCTGGCGCAGACGCTCGGCAAGGAGCAGCCCTTCTACG 3901

Qy       625 CCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCG 684
Db      3902 GGCTCCAGGCGCGTGGCGTGGACGGCAAGCTGCCTCCGGCCGACAGCATCGAGGAGATGG 3961

Qy       685 CGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTG 744
Db      3962 CGTCCATCTACCTGGAGAGCATCCGTCAGGTGCAGCCCCACGGGCCCTACCTGCTCGGCG 4021

Qy       745 GGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGG 804
Db      4022 GCTACTCCGGCGGCGGCGTGGTCGCCTACAAAATAGCGCAGCGGCTGCACGCGCTGGGCG 4081

Qy       805 AGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGG 864
Db      4082 AGCCGGTGGCGCTGGTCGCGTTCCTGGACACCTTCCACCCGTCCACCCAGGAGCGACGCC 4141

Qy       865 AGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCA 924
Db      4142 AGTCCCTCCATGAGCGCCTGCGGGAGCTGCGCCGGGAGGGCGCGGGCAGCTACGTCAGCC 4201

Qy       925 G 925
Db      4202 G 4202
```

RESULT 3
ACL64689/c

ID ACL64689 standard; DNA; 17727 BP.
XX
AC ACL64689;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus DNA fragment, seq id 1152.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.
XX
PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 1; SEQ ID NO 1152; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC set of 1849 contig and singleton sequences comprising coding sequences,
CC DNA replication elements, promoters and other regulatory elements from
CC the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO
XX
SQ Sequence 17727 BP; 2758 A; 6082 C; 6141 G; 2745 T; 0 U; 1 Other;

Query Match 12.0%; Score 120.2; DB 14; Length 17727;
Best Local Similarity 52.2%; Pred. No. 1.6e-08;
Matches 314; Conservative 0; Mismatches 283; Indels 4; Gaps 2;

```
Qy       326 CGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGC 385
Db     14123 CGCTGGGCGCGGACCTGACGCTGGCCACGCTGTTCGAGGCTCCCACGCTGGAGCAGTGCG 14064

Qy       386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTC-CGGCGCCTGATCCGCGCGCGGGGGACG 444
Db     14063 CCGCGCTGGTGCGTGAGGCCGCCGGGATTCCCTTCACGCCGGACGCGGGTCCGGGTGACG 14004

Qy       445 CACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGC 504
Db     14003 CGGCGCCGGCTGCGCAGGCCCCGGGCGTCAAGGCGGTGCCGAAG---AGCTGGTCTCCGC 13947

Qy       505 TCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTG 564
Db     13946 TGGTCACCATCCAGAAGGGCGGCAATACCCTCCCCTTCTTCTGCGTGCACGGCGCGGGCG 13887

Qy       565 GGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCG 624
Db     13886 GCAACGTCCTCAACTTCCGCGAGCTGGCGCAGACGCTCGGCAAGGAGCAGCCCTTCTACG 13827

Qy       625 CCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCG 684
```

```
                ||||||| || ||||| |||||          ||   |   |||||||||||||| |
Db      13826 GGCTCCAGGCGCGTGGCGTGGACGGCAAGCTGCCTCCGGCCGACAGCATCGAGGAGATGG 13767

Qy        685 CGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTG 744
              || ||   ||| | |   || | || |||      ||    ||||| ||| |  | || |
Db      13766 CGTCCATCTACCTGGAGAGCATCCGTCAGGTGCAGCCCCACGGGCCCTACCTGCTCGGCG 13707

Qy        745 GGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGG 804
              |   |||  ||||||| |  |||||| | | || || | || ||| | | |||  || |
Db      13706 GCTACTCCGGCGGCGGCGTGGTCGCCTACAAAATAGCGCAGCGGCTGCACGCGCTGGGCG 13647

Qy        805 AGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGG 864
              ||| ||  | | | | |   | || |||||||  | | ||  | |           |
Db      13646 AGCCGGTGGCGCTGGTCGCGTTCCTGGACACCTTCCACCCGTCCACCCAGGAGCGACGCC 13587

Qy        865 AGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCA 924
              || | |     ||   |   | |   ||||  |    |  |  || || || | ||| |
Db      13586 AGTCCCTCCATGAGCGCCTGCGGGAGCTGCGCCGGGAGGGCGCGGGCAGCTACGTCAGCC 13527

Qy        925 G 925
              |
Db      13526 G 13526
```

RESULT 4
ACA25936
ID ACA25936 standard; DNA; 1121 BP.
XX
AC ACA25936;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #7593.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Burkholderia mallei.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU22066.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 13806; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1121 BP; 158 A; 386 C; 409 G; 168 T; 0 U; 0 Other;

 Query Match 11.0%; Score 110; DB 8; Length 1121;
 Best Local Similarity 48.8%; Pred. No. 7e-07;
 Matches 296; Conservative 0; Mismatches 310; Indels 0; Gaps 0;

```
Qy        308 CGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGC 367
              ||  || | ||  |   ||| | |   || | |   |   ||     || ||  ||  |
Db         63 CGACGCGCTGGGCGTGACCGTACCGGTCGAGTCGATCCTCGCGAGTCCGACCATCGTCGA 122

Qy        368 CCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGA 427
              |  || |  | ||  |           | ||  ||   ||| ||| |   | |||||  |
Db        123 GCTTGCCGGATACGTCGAGCAATTGCGCTCGCGGGACGTCCGCGACGGCGCTGCGCCCGT 182

Qy        428 TCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGG 487
              ||  ||  |       |   |  || ||  |   |         ||||| | ||  |||
Db        183 GTCGATCGGCGAAACACCGGCGGACGCGGATGCTCGAGCTCAGGCGCAGGCGGATACGGT 242

Qy        488 GCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTG 547
              || || |     |  | ||||| |           |||  |  ||  | |    ||||||
Db        243 GCACACCGATTGCCTGATCGTCATTCAAGCAGGCGGCGCCGAACAAGCGCCGGTGTTCTG 302

Qy        548 GGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCAC 607
              | |  ||||||   ||   ||    | || ||  | | ||  |   |  ||| || | |
Db        303 CATCCCGGGCGCGGGGGGCAGCGTCGCGTCGTTCGTTGCGCTTGCGAGCATGCTGCGCGC 362

Qy        608 GGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCAC 667
              ||    ||| | | || ||| ||| | || ||  | ||||| || |    |||   |
Db        363 CGACATACCGGTATACGGCTTGCAGCCTCGCGGGCTGGACGGCCTGGGGCCGCCGGACCG 422

Qy        668 GAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGG 727
              | |||| | |   ||| ||  ||||       | | | | |    | |||   ||
Db        423 ATCCGTCGAAGCGGCTGCGCGCCGGTACGCGCGGGCCATTCTGGATGCCACCCCGCCCGG 482

Qy        728 GCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCA 787
              |||       | | |  || |  || | |||||||   |||||  ||||||  || ||||
Db        483 GCCGCCGCGCATCGTCGGCCACTCGTTCGGCGGCTGGATCGCGCTCGAGACAGCGCGGCT 542

Qy        788 GCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGT 847
              |||   |      ||   ||||  | ||| ||| ||| |||||||   |  ||||  ||
Db        543 GCTGGACGGCATGGGAACGCGCTGCGCCCCGCTCGTCCTGCTCGATTCGAATCCACCGCC 602

Qy        848 CGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCAT 907
              ||||   |  ||   |   ||||   |||    |||  | | |   |||| |   | |
Db        603 CGCGTCGCAGGCCTGGCGCGCGCCTTCCGAGGCAGACATGCTGCGCACGCTCGTCGGCCT 662

Qy        908 GGCCAA 913
              |  | |
Db        663 GCTCGA 668
```

```
RESULT 5
ADJ72363
ID   ADJ72363 standard; DNA; 90597 BP.
XX
AC   ADJ72363;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Streptomyces roseosporus daptomycin biosynthetic gene cluster DNA.
XX
KW   ds; gene; antibacterial; gene therapy;
KW   daptomycin biosynthesis gene cluster;
KW   daptomycin non-ribosomal peptide synthetase; DptBC;
KW   gram-positive bacterial infection.
XX
OS   Streptomyces roseosporus.
XX
PN   WO2003014297-A2.
XX
PD   20-FEB-2003.
XX
PF   31-JUL-2002; 2002WO-US024310.
XX
PR   06-AUG-2001; 2001US-0310385P.
PR   17-OCT-2001; 2001WO-US032354.
PR   10-MAY-2002; 2002US-0379866P.
XX
PA   (CUBI-) CUBIST PHARM INC.
XX
PI   Miao VPW,  Brian P,  Baltz RH,  Coeffet-Legal MF;
XX
DR   WPI; 2003-268192/26.
XX
PT   New isolated nucleic acid molecule encoding a daptomycin non-ribosomal
PT   peptide synthetase, useful for treatment of a gram-positive bacterial
PT   infection of skeletal muscle, skin, bloodstream, kidneys, heart, lung and
PT   bone.
XX
PS   Disclosure; SEQ ID NO 1; 292pp; English.
XX
CC   The invention relates to new isolated nucleic acid (NA) molecules from
CC   the Streptomyces roseosporus daptomycin biosynthesis gene cluster,
CC   especially a daptomycin non-ribosomal peptide synthetase (NRPS) or its
CC   subunit, where the (NA) molecule encodes DptBC, and is not pRHB159. The
CC   methods and compositions of the present invention are useful for
CC   treatment of a gram-positive bacterial infection of any organ or tissue
CC   in the body, including skeletal muscle, skin, bloodstream, kidneys,
CC   heart, lung and bone. This sequence represents the daptomycin
CC   biosynthesis gene cluster sequence.
XX
SQ   Sequence 90597 BP; 12671 A; 32312 C; 31569 G; 14045 T; 0 U; 0 Other;

  Query Match             10.8%;  Score 107.8;  DB 10;  Length 90597;
  Best Local Similarity   47.6%;  Pred. No. 6.1e-07;
  Matches  445;  Conservative    0;  Mismatches  482;  Indels    8;  Gaps    4;

Qy         42 GACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCC 101
              ||||    ||| ||| ||  ||||  |   |||| ||  |    ||  ||    | ||| |
Db      83946 GACGGGCTCGGCGACCAGCGGCTCGTCGCGTACGTCACCGGCGGCC-CCGGCACACCGTC 84004

Qy        102 TTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACC 161
                |   ||  | | ||  |  |    ||   ||   |  |||       |  || |
Db      84005 GGCCGCCGCGCTGCGCGCCCACCTGGCGTCCCGGCTGCCCCGGCACATGGTGCCCGGTGA 84064

Qy        162 CATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCA 221
              | |||   |       |  |   | || |||   ||  | |       |      |||||||
Db      84065 CGTCCTCACCCTGGACGCCCTGCCGCTCACCGCCAACGGCAAGGTGGACCGCACGGCGCT 84124

Qy        222 GCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCG 281
              ||||||| || |  | | |  | || | ||  | |     |  |  |||| | |   ||
Db      84125 GCCCGGCCCCGGCACCCAGACCGCCGCCCCCGGGCGCGCACCCCAGTCGCCGCAGGAACG 84184

Qy        282 GCGGC--CTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGG 339
              |  ||    |  |     || || | | | ||||  |    | || ||     |   ||
Db      84185 GGTGCTGTGCGCCTTGTTCGCCGACGTGCTCGGCCGGGAGACCGTCGGCGTGGACGAGGG 84244

Qy        340 CCTCCGCAGCCCCGGCGG--CCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
                ||  |  ||  |||||   |||   |  | || |   |  |||||||   |    | |
Db      84245 GTTCTTCGACCTGGGCGGTCACTCGCTGCTCGCCACTCGCCTCGCGGCCCGGGTCCGCGC 84304

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
              ||||  |  |   |||    | | ||     | |  || || |    | |   ||  |
Db      84305 GGCGCTGGGCGTGGAGATCTCCGTGCGCACCCTGTTCGAGGCGCCGACCCCTGCCCTGCT 84364

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
                |          | ||  |  ||  ||| |  | ||||   ||||   ||  |
Db      84365 CGCGTCGGCGTGCACGGCGGACGCCGCGGCGTACGACCCGTTCGAGACGGTGCTGCCGCT 84424

Qy        518 CCCGCGCGGCTCACGCGGGAGCTC---GTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCA 574
              | ||||      || ||   |  ||||| || |||| ||     || ||
Db      84425 GCGGCGCACGGGCAGCCGGCCACCGCTGTTCTGCGTCCACGCCGGAATGGGCCTGAGCTG 84484

Qy        575 GCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGC 634
              | ||  |     ||||      |   || | | ||||    ||| | | ||   | |||||
Db      84485 GGCGTACGCCGGCCTGCTCAGCCATCTGGACGCGGACGTGCCGGTTTACGGACTGCAGGC 84544

Qy        635 CCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTA 694
              |||| |  ||  || || || |  ||     | |||| |||||||||| || |   | ||
Db      84545 CCGGAGGCTCACCGCGCCCGGCGGGCTGCCCGGGAGCGTCGAGGAGATGGCTGAGGACTA 84604

Qy        695 CGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTC 754
              || ||      | ||||  |   | |||   ||||| |||    ||   ||   |||||
Db      84605 CGCCGGTGAGATCCGGCGCCTGTGCCCGGATGGGCCGTACCGGCTGCTCGGCTGGTCCTT 84664

Qy        755 CGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTC 814
              ||||||||  |||||  ||   | ||     |  || |    ||| || |   ||||
Db      84665 CGGCGGCACGGTCGCCCACGCCGTCGCGACCCGCCTGCAACAGCAGGGCCACACCGTCGA 84724

Qy        815 CCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGG 874
              |  || | | |  |||||| |||| ||||    || || ||   |  | |    |  ||
Db      84725 ACTCCTCGCCGTCCTCGACGCCTACCCCGTCACCGGGGCCCGGCCCGACGCCGAGGTGGA 84784

Qy        875 CGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCG 934
              |||      |   ||       |||  |  |  |  |||  | ||  |  ||| ||
Db      84785 CGAACAGCGCATCGTCGCCGACTACCTCGCCCAGCTCGGTTCCCCCGTCGCCCCCGAGCG 84844

Qy        935 CGACGGAGACACGGCGATCAAGCCCGACGACCTCG 969
              |  ||  | |   |||       || ||   |||||
Db      84845 CCTCGAGGGCGACGCGTGGCTCCCGGAGTTCCTCG 84879


RESULT 6
ABQ78872
ID   ABQ78872 standard; DNA; 90600 BP.
XX
AC   ABQ78872;
XX
DT   23-OCT-2002  (first entry)
XX
DE   S. roseosporus daptomycin biosynthetic gene cluster 90kb region.
XX
KW   Daptomycin biosynthetic gene cluster; thioesterase; antibacterial;
KW   fungicide; virucide; antiparasitic; immunomodulator; antilipemic;
KW   cytostatic; gene therapy; antimitotic; immunomodulatory; siderophore;
KW   anti-cholesterolemic; agrochemical; gene; ds.
XX
OS   Streptomyces roseosporus.
XX
PN   WO200259322-A2.
XX
PD   01-AUG-2002.
XX
PF   17-OCT-2001; 2001WO-US032354.
XX
```

```
PR   17-OCT-2000; 2000US-0240879P.
PR   28-FEB-2001; 2001US-0272207P.
PR   06-AUG-2001; 2001US-0310385P.
XX
PA   (MIAO/) MIAO V P W.
PA   (BRIA/) BRIAN P.
PA   (BALT/) BALTZ R H.
PA   (SILV/) SILVA C J.
XX
PI   Miao VPW,  Brian P,  Baltz RH,  Silva CJ;
XX
DR   WPI; 2002-599794/64.
XX
PT   Isolated nucleic acid molecule from a bacterial daptomycin biosynthetic
PT   gene cluster encoding a thioesterase or thioesterase domain, useful for
PT   generating novel linear and cyclic peptides, and products in a cell.
XX
PS   Claim 7; Page 142-165; 227pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule
CC   comprising a sequence that encodes a thioesterase or thioesterase domain,
CC   derived from a bacterial daptomycin biosynthetic gene cluster. The
CC   proteins of the invention have antibacterial, fungicide, virucide,
CC   antiparasitic, immunomodulator, antilipemic, and cytostatic activity. The
CC   polynucleotides may have a use in gene therapy. The compositions and
CC   methods of the present invention are useful for generating novel linear
CC   and cyclic peptides and improving yield of a product in a cell expressing
CC   an daptomycin non-ribosomal peptide synthetase (NRPS) to be used as new
CC   compounds or in producing new compounds, such as antibiotics,
CC   antifungals, antivirals, antiparasitics, antimitotics, antitumour agents,
CC   immunomodulatory agents, anti-cholesterolemic agents, siderophores,
CC   agrochemicals and cytostatics. The sequence represents the 90kb region of
CC   the S. roseosporus daptomycin biosynthetic gene cluster
XX
SQ   Sequence 90600 BP; 12671 A; 32312 C; 31571 G; 14046 T; 0 U; 0 Other;

  Query Match             10.8%;  Score 107.8;  DB 6;  Length 90600;
  Best Local Similarity   47.6%;  Pred. No. 6.1e-07;
  Matches  445;  Conservative    0;  Mismatches  482;  Indels    8;  Gaps    4;

Qy         42 GACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCC 101
Db      83948 GACGGGCTCGGCGACCAGCGGCTCGTCGCGTACGTCACCGGCGGCC-CCGGCACACCGTC 84006

Qy        102 TTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACC 161
Db      84007 GGCCGCCGCGCTGCGCGCCCACCTGGCGTCCCGGCTGCCCCGGCACATGGTGCCCGGTGA 84066

Qy        162 CATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCA 221
Db      84067 CGTCCTCACCCTGGACGCCCTGCCGCTCACCGCCAACGGCAAGGTGGACCGCACGGCGCT 84126

Qy        222 GCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCG 281
Db      84127 GCCCGGCCCCGGCACCCAGACCGCCGCCCCCGGGCGCGCACCCCAGTCGCCGCAGGAACG 84186

Qy        282 GCGGC--CTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGG 339
Db      84187 GGTGCTGTGCGCCTTGTTCGCCGACGTGCTCGGCCGGGAGACCGTCGGCGTGGACGAGGG 84246

Qy        340 CCTCCGCAGCCCCGGCGG--CCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
Db      84247 GTTCTTCGACCTGGGCGGTCACTCGCTGCTCGCCACTCGCCTCGCGGCCCGGGTCCGCGC 84306

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
Db      84307 GGCGCTGGGCGTGGAGATCTCCGTGCGCACCCTGTTCGAGGCGCCGACCCCTGCCCTGCT 84366

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
Db      84367 CGCGTCGGCGTGCACGGCGGACGCCGCGGCGTACGACCCGTTCGAGACGGTGCTGCCGCT 84426

Qy        518 CCCGCGCGGCTCACGCGGGAGCTC---GTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCA 574
Db      84427 GCGGCGCACGGGCAGCCGGCCACCGCTGTTCTGCGTCCACGCCGGAATGGGCCTGAGCTG 84486

Qy        575 GCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGC 634
Db      84487 GGCGTACGCCGGCCTGCTCAGCCATCTGGACGCGGACGTGCCGGTTTACGGACTGCAGGC 84546

Qy        635 CCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTA 694
Db      84547 CCGGAGGCTCACCGCGCCCGGCGGGCTGCCCGGGAGCGTCGAGGAGATGGCTGAGGACTA 84606

Qy        695 CGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTC 754
Db      84607 CGCCGGTGAGATCCGGCGCCTGTGCCCGGATGGGCCGTACCGGCTGCTCGGCTGGTCCTT 84666

Qy        755 CGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTC 814
Db      84667 CGGCGGCACGGTCGCCCACGCCGTCGCGACCCGCCTGCAACAGCAGGGCCACACCGTCGA 84726

Qy        815 CCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGG 874
Db      84727 ACTCCTCGCCGTCCTCGACGCCTACCCCGTCACCGGGGCCCGGCCCGACGCCGAGGTGGA 84786

Qy        875 CGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCG 934
Db      84787 CGAACAGCGCATCGTCGCCGACTACCTCGCCCAGCTCGGTTCCCCCGTCGCCCCCGAGCG 84846

Qy        935 CGACGGAGACACGGCGATCAAGCCCGACGACCTCG 969
Db      84847 CCTCGAGGGCGACGCGTGGCTCCCGGAGTTCCTCG 84881


RESULT 7
ADO13830/c
ID   ADO13830 standard; DNA; 71138 BP.
XX
AC   ADO13830;
XX
DT   15-JUL-2004  (first entry)
XX
DE   dbv gene cluster.
XX
KW   ss; gene; dbv; gene cluster; Actinomadura sp.; biosynthesis;
KW   glycopeptide; A40926; 4-hydroxy-phenylglycine; 3;
KW   5-dihydroxy-phenylglycine; heptapeptide skeleton; chlorination;
KW   aromatic residue; beta-hydroxylation; tyrosine; cross-linking;
KW   N-acetyl glucuroamine; mannosyl; export; antibiotic.
XX
OS   Actinomadura sp. ATCC 39727.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(40. .1140)
FT                   /*tag=  a
FT                   /label= ORF1
FT                   /transl_except= pos:1138. .1140,  aa:Met
FT   CDS             complement(1259. .2329)
FT                   /*tag=  b
FT                   /label= ORF2
FT   CDS             complement(2558. .5161)
FT                   /*tag=  c
FT                   /label= ORF3
FT                   /transl_except= pos:5159. .5161, aa:Met
FT   CDS             complement(5266. .6231)
FT                   /*tag=  d
FT                   /transl_except= pos:6229. .6231, aa:Met
FT                   /label= ORF4
FT   CDS             7183. .8292
FT                   /*tag=  e
FT                   /label= ORF5
FT                   /transl_except= pos:7183. .7185, aa:Met
FT   CDS             8320. .8973
FT                   /*tag=  f
FT                   /label= ORF6
```

```
FT   CDS             9069. .9659
FT                   /*tag= g
FT                   /label= ORF7
FT   CDS             complement(9708. .10667)
FT                   /*tag= h
FT                   /label= ORF8
FT   CDS             complement(10670. .11896)
FT                   /*tag= i
FT                   /label= ORF9
FT   CDS             complement(11950. .13419)
FT                   /*tag= j
FT                   /label= ORF10
FT                   /transl_except= pos:13417. .13419, aa:Met
FT   CDS             complement(13479. .14741)
FT                   /*tag= k
FT                   /transl_except= pos:14739. .14741, aa:Met
FT                   /label= ORF11
FT   CDS             complement(14823. .16019)
FT                   /*tag= l
FT                   /label= ORF12
FT                   /transl_except= pos:16017. .16019, aa:Met
FT   CDS             complement(16009. .17163)
FT                   /*tag= m
FT                   /label= ORF13
FT   CDS             complement(17185. .18366)
FT                   /*tag= n
FT                   /label= ORF14
FT                   /transl_except= pos:18364. .18366, aa:Met
FT   CDS             complement(18462. .18671)
FT                   /*tag= o
FT                   /label= ORF15
FT   CDS             complement(18668. .24259)
FT                   /*tag= p
FT                   /label= ORF16
FT                   /transl_except= pos:24257. .24259, aa:Met
FT   CDS             complement(24278. .36529)
FT                   /*tag= q
FT                   /label= ORF17
FT   CDS             complement(36760. .39021)
FT                   /*tag= r
FT                   /label= ORF18
FT   CDS             complement(39153. .39851)
FT                   /*tag= s
FT                   /label= ORF19
FT   CDS             complement(40125. .41732)
FT                   /*tag= t
FT                   /label= ORF20
FT   CDS             complement(41772. .42584)
FT                   /*tag= u
FT                   /label= ORF21
FT   CDS             complement(42868. .44130)
FT                   /*tag= v
FT                   /label= ORF22
FT                   /transl_except= pos:44128. .44130, aa:Met
FT   CDS             complement(44226. .46355)
FT                   /*tag= w
FT                   /label= ORF23
FT   CDS             46632. .48578
FT                   /*tag= x
FT                   /label= ORF24
FT   CDS             48575. .54868
FT                   /*tag= y
FT                   /label= ORF25
FT   CDS             54865. .58056
FT                   /*tag= z
FT                   /label= ORF26
FT                   /transl_except= pos:54865. .54867, aa:Met
FT   CDS             58152. .58985
FT                   /*tag= aa
FT                   /label= ORF27
FT   CDS             59046. .60641
FT                   /*tag= ab
FT                   /label= ORF28
FT   CDS             complement(60874. .62445)
FT                   /*tag= ac
FT                   /label= ORF29
FT                   /transl_except= pos:62443. .62445, aa:Met
FT   CDS             62887. .63312
FT                   /*tag= ad
FT                   /label= ORF30
FT   CDS             63469. .64587
FT                   /*tag= ae
FT                   /label= ORF31
FT   CDS             64599. .65240
FT                   /*tag= af
FT                   /label= ORF32
FT   CDS             65237. .66541
FT                   /*tag= ag
FT                   /label= ORF33
FT                   /transl_except= pos:65237. .65239, aa:Met
FT   CDS             66538. .67335
FT                   /*tag= ah
FT                   /label= ORF34
FT   CDS             67332. .68618
FT                   /*tag= ai
FT                   /label= ORF35
FT   CDS             complement(68685. .69440)
FT                   /*tag= aj
FT                   /label= ORF36
FT   CDS             69608. .70894
FT                   /*tag= ak
FT                   /label= ORF37
FT   misc_feature    71065. .71138
FT                   /*tag= al
FT                   /note= "Remnant of attL site"
XX
PN   EP1413626-A1.
XX
PD   28-APR-2004.
XX
PF   23-OCT-2002; 2002EP-00023597.
XX
PR   23-OCT-2002; 2002EP-00023597.
XX
PA   (VICU-) VICURON PHARM INC.
XX
PI   Donadio S,  Sosio M,  Beltrametti F;
XX
DR   WPI; 2004-332813/31.
DR   P-PSDB; ADO13831, ADO13832, ADO13833, ADO13834, ADO13835, ADO13836,
DR   ADO13837, ADO13838, ADO13839, ADO13840, ADO13841, ADO13842, ADO13843,
DR   ADO13844, ADO13845, ADO13846, ADO13847, ADO13848, ADO13849, ADO13850,
DR   ADO13851, ADO13852, ADO13853, ADO13854, ADO13855, ADO13856, ADO13857,
DR   ADO13858, ADO13859, ADO13860, ADO13861, ADO13862, ADO13863, ADO13864,
DR   ADO13865, ADO13866, ADO13867.
XX
PT   Novel isolated polypeptide A40926, comprising polypeptide sequence or
PT   open reading frame polypeptide involved in biosynthetic pathway of
PT   A409626, useful for synthesizing glycopeptide antibiotic A409626.
XX
PS   Claim 1; Page 25-63; 164pp; English.
XX
CC   This sequence represents the dbv gene cluster from Nonomuria sp.
CC   ATCC39727 (formerly Actinomadura sp. ATCC39727). The 37 proteins encoded
CC   by this sequence are involved in the biosynthesis of the glycopeptide
CC   A40926. dbv ORF 1, 2, 5, and 37 encode polypeptides required for the
CC   synthesis of 4-hydroxy-phenylglycine residues of A40926. dbv ORF 30-34
CC   and 37 encode polypeptides required for the synthesis of 3, 5-dihydroxy-
CC   phenylglycine residues of A40926. dbv ORF 16, 17, 25, 26 and 36 encode
CC   polypeptides required for the synthesis of heptapeptide skeleton of
CC   A40926. dbv ORF 10 encodes polypeptide required for the chlorination of
CC   the aromatic residues of A40926. dbv ORF 28 encodes a polypeptide
CC   required for the beta-hydroxylation of the tyrosine residue of amino acid
CC   6 of A40926. dbv ORF 11-14 encode polypeptides required for cross-linking
CC   of the aromatic residues of amino acids at positions 2 and 4, 4 and 6, 1
CC   and 3, and 5 and 7 of A40926. dbv ORF 9, 23 and 29 encode polypeptides
```

```
CC   required for addition and formation of N-acetyl glucuroamine residue of
CC   A40926. dbv ORF 20 or 27 encode polypeptides required for the attachment
CC   of mannosyl residues or N-methylation of A40926. dbv ORF 7, 18, 19, 24
CC   and 35 encode polypeptides required for the export of A40926 or its
CC   precursors. dbv ORF 3, 4, 6, and 22 encode polypeptides required for
CC   regulating the expression of one or more genes of the dbv gene cluster.
CC   The dbv gene cluster and the proteins encoded by it are useful for
CC   producing glycopeptide antibiotic A40926 or its precursor.
XX
SQ   Sequence 71138 BP; 10493 A; 24958 C; 25226 G; 10461 T; 0 U; 0 Other;

  Query Match             10.8%;  Score 107.6;  DB 12;  Length 71138;
  Best Local Similarity   48.8%;  Pred. No. 6.8e-07;
  Matches  327;  Conservative    0;  Mismatches  334;  Indels    9;  Gaps    1;

Qy        218 CGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGC 277
Db      19718 CGAGGCCGTGCCCCCGGTGGCCGCCGACCTGGCCGAGGGCGGTCCGCTCGGCGCGCTCGT 19659

Qy        278 CCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGC 337
Db      19658 GCTCGACGACCGGCTGCGCCCCGCCGTCGCCGTCGGCGAGCTCTACCTCACCGGCGCGGC 19599

Qy        338 GGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
Db      19598 GGTGGACGCGGAGCCGGGCGACCGGACCCTGGCCTGCCCCTTCGGGGCCACGGGGCGGCG 19539

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
Db      19538 CATGCTGCCCACGGGCCTGCTCGCCCGCTGGACAGCCGGGGGCACCCTGGTCGTCGTCGG 19479

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
Db      19478 CGAGCGGCGGGGATCGAGCGGCTCGGTGAAGACCGGCACGGGCGACTTCGAGGTGCTGCT 19419

Qy        518 CCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCC 577
Db      19418 GCCGCTGCGAGCCGGCGGTAACCGTCCGCCGCTGTACTGCGTCCATGCGAGCGGAGGGCT 19359

Qy        578 GCTCTATCCCCTGTCTGCCGCGCTCGG---------CACGGATTACCCGTTCTTCGCCTT 628
Db      19358 GAGCTGGAACTACGCGCCCCTGCTGCGGAGCCTGCCCCCCAACCAGCCGGTCTACGGCGT 19299

Qy        629 CCAGGCCCGGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGC 688
Db      19298 GCAGGCGCGCGGCCTGGCTCGTACGGAGCCGCTGGCGGCCGGCGTCGAGGAGATGGCGGC 19239

Qy        689 CCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCT 748
Db      19238 CGATTACGTCGAGCAGATCCGCGCCGTGCAGCCGACCGGGCCGTACCACCTCCTCGGGTG 19179

Qy        749 GTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCG 808
Db      19178 GTCACTGGGCGGGCGGATCGCGCAGGAGATGGCCAGGGTGCTGGAGGAGCAAGGGGAGCA 19119

Qy        809 CGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGAC 868
Db      19118 GGTCGGCCTGCTCGCCCTGCTCGACGCCTATCCCACCGACGTGGGCAGGCTCCGGCGCCC 19059

Qy        869 GCCGGGCGAC 878
Db      19058 GCGCGGCGAC 19049


RESULT 8
ADS14647
ID   ADS14647 standard; DNA; 7059 BP.
XX
AC   ADS14647;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Pseudomonas aeruginosa quorum sensing controlled gene PA3327, SEQ ID 202.
XX
KW   Cytostatic; Anti-HIV; Antimicrobial; Immunosuppressive;
KW   quorum sensing signalling; bacterium; quorum sensing controlled gene;
KW   biofilm-associated disease; cystic fibrosis; AIDS; middle ear infection;
KW   acne; periodontal disease; gene; ds; PA3327.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO2004083385-A2.
XX
PD   30-SEP-2004.
XX
PF   11-MAR-2004; 2004WO-US007467.
XX
PR   14-MAR-2003; 2003US-00389647.
XX
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Greenberg EP,  Schuster M,  Lostroh C;
XX
DR   WPI; 2004-709932/69.
XX
PT   Identifying a modulator of quorum sensing signaling in bacteria, useful
PT   for treating a biofilm-associated disorder, comprises contacting the cell
PT   with a quorum sensing signal molecule in the presence and absence of a
PT   test compound.
XX
PS   Disclosure; SEQ ID NO 202; 233pp; English.
XX
CC   The present invention relates to a method for identifying a modulator of
CC   quorum sensing signalling in bacteria. The method comprises: providing a
CC   cell that comprises a quorum sensing controlled gene (ADS14446-ADS14798),
CC   where the cell is responsive to a quorum sensing signal molecule such
CC   that a detectable signal is generated; contacting the cell with a quorum
CC   sensing signal molecule in the presence and absence of a test compound;
CC   and detecting a change in the detectable signal. The method and modulator
CC   identified by the method are useful for treating a biofilm-associated
CC   disease or disorder, e.g. cystic fibrosis, AIDS, middle ear infections,
CC   acne, periodontal disease, catheter-associated infections, and medical
CC   device-associated infections. Note: The sequence data for this patent was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7059 BP; 1036 A; 2660 C; 2302 G; 1061 T; 0 U; 0 Other;

  Query Match             10.6%;  Score 106.4;  DB 13;  Length 7059;
  Best Local Similarity   50.8%;  Pred. No. 1.6e-06;
  Matches  308;  Conservative    0;  Mismatches  291;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
Db       5956 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6015

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
Db       6016 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6075

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
Db       6076 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6131

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
Db       6132 GACGCCGACCGTGGAGGGCCTGGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6191

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
Db       6192 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6251

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
Db       6252 GCTCGGCGGCCACGTGCTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6311
```

```
Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |   | |||  |  ||  | ||||
Db       6312 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6371

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              ||| ||||| |||  |||| ||||   | | || | |      |||  ||| || ||
Db       6372 GGACATCGCCGCCAGTTACCTCGCGGCCATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6431

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | |||||| || |||||| |||||||||||||| || | | |
Db       6432 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6491

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |  | ||   ||| |||  ||| || |||||||| ||    |||  | |
Db       6492 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6551

Qy        854 CCTCGC 859
              || |||
Db       6552 CCACGC 6557
```

RESULT 9
AAF81361
ID AAF81361 standard; DNA; 7110 BP.
XX
AC AAF81361;
XX
DT 04-JUN-2001 (first entry)
XX
DE Quorum sensing controlled gene qsc126 ORF.
XX
KW Quorum sensing; antibacterial; bacterial signalling;
KW opportunistic pathogen; immunocompromised; burn; cystic fibrosis;
KW immunosuppressive therapy; AIDS; ss.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200118248-A2.
XX
PD 15-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-US024141.
XX
PR 03-SEP-1999; 99US-0153022P.
XX
PA (IOWA) UNIV IOWA RES FOUND.
PA (QUOR-) QUORUM SCI INC.
XX
PI Whiteley M, Lee KM, Greenberg EP, Muh U;
XX
DR WPI; 2001-265973/27.
XX
PT Identifying modulators of quorum sensing signaling in Pseudomonas
PT aeruginosa bacteria, useful for treating infections in immunocompromized
PT patients.
XX
PS Claim 46; Page 103-5; 115pp; English.
XX
CC Bacteria signal to one another to coordinate expression of specific genes
CC in a cell density dependent fashion. This "bacterial signalling" is
CC called "quorum sensing and response". Quorum sensing allows a bacterial
CC species to sense its own number and regulate gene expression according to
CC population density. The present sequence is an open reading frame (ORF)
CC of a Pseudomonas aeruginosa quorum sensing controlled gene. Inhibitors of
CC quorum sensing signalling renders a bacterial population more susceptible
CC to treatment. The present invention relates to a method for identifying
CC modulators of quorum sensing signalling in Pseudomonas aeruginosa
CC bacteria. Modulators of quorum signaling may be used to treat P.
CC aeruginosa infections. P. aeruginosa is an opportunistic pathogen of
CC immunocompromised individuals (burn patients, cystic fibrosis patients,
CC patients undergoing immunosuppressive therapy and patients with AIDS)
XX
SQ Sequence 7110 BP; 1045 A; 2670 C; 2323 G; 1072 T; 0 U; 0 Other;

```
  Query Match             10.6%;  Score 106.4;  DB 4;  Length 7110;
  Best Local Similarity   50.8%;  Pred. No. 1.6e-06;
  Matches  308;  Conservative    0;  Mismatches  291;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||             ||   ||
Db       6007 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6066

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              | ||| |||||  |  |  |  |  || | ||  |  |   ||| |||      |||  |
Db       6067 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6126

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
                | |      ||   |  | |  ||||     |||    |||  | | |  | |  ||
Db       6127 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6182

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
              | |    |  |   ||  || |      |  | | | | |  | ||| || | |
Db       6183 GACGCCGACCGTGGAGGGCCTGGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6242

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              ||||  ||| ||    |||    |  |||||  |||  |     ||||   || |||
Db       6243 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6302

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
                  || | || ||       |   || |||     ||| ||     ||||  | ||
Db       6303 GCTCGGCGGCCACGTGCTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6362

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |   | |||  |  ||  | ||||
Db       6363 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6422

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              ||| ||||| |||  |||| ||||   | | || | |      |||  ||| || ||
Db       6423 GGACATCGCCGCCAGTTACCTCGCGGCCATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6482

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| |||||||||||||| || | | |
Db       6483 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6542

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |  | ||   ||| |||  ||| || |||||||| ||    |||  | |
Db       6543 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6602

Qy        854 CCTCGC 859
              || |||
Db       6603 CCACGC 6608
```

RESULT 10
AAD54224
ID AAD54224 standard; DNA; 5862 BP.
XX
AC AAD54224;
XX
DT 17-JUN-2003 (first entry)
XX
DE Streptomyces platensis subspecies rosaceus dorrigocin ORF7 DNA.
XX
KW Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
KW ds.
XX
OS Streptomyces platensis.
XX
FH Key Location/Qualifiers
FT CDS 1. .5862
FT /*tag= a
FT /product= "ORF7 protein"
FT /note= "No start codon"
FT /partial
XX
PN WO200288176-A2.

```
XX
PD   07-NOV-2002.
XX
PF   26-APR-2002; 2002WO-CA000591.
XX
PR   26-APR-2001; 2001US-0286346P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazopoulos E,  Staffa A,  Yang X;
XX
DR   WPI; 2003-201222/19.
DR   P-PSDB; AAE35490.
XX
PT   Novel isolated or purified polypeptide involved in biosynthesis of
PT   polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
PT   dorrigocin or lactimidomycin.
XX
PS   Claim 6; Page 188-191; 312pp; English.
XX
CC   The invention relates to novel proteins involved in the biosynthesis of
CC   polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC   microorganisms. Sequences of the invention allow direct manipulation of
CC   dorrigocin, lactimidomycin and related chemical structures via chemical
CC   engineering of the enzymes involved in the biosynthesis of dorrigocin and
CC   lactimidomycin. They are useful for introducing chemical handles into
CC   normally inert positions that permit subsequence chemical modifications
CC   and facilitate the development of polyketides. The genes and proteins of
CC   the invention can also be used to generate a focused library of analogues
CC   around a polyketide lead candidate to fine-tune the compound for optimal
CC   properties. They are useful for generating antibodies specific for the
CC   polyketide biosynthesis. The present sequence is Streptomyces platensis
CC   subspecies rosaceus DORR ORF7 DNA
XX
SQ   Sequence 5862 BP; 878 A; 2254 C; 1963 G; 767 T; 0 U; 0 Other;

  Query Match             10.3%;  Score 102.8;  DB 10;  Length 5862;
  Best Local Similarity   52.4%;  Pred. No. 5.1e-06;
  Matches  302;  Conservative    0;  Mismatches  262;  Indels   12;  Gaps    3;

Qy        260 GCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGC 319
              || ||   |||||  ||| | | ||  |  ||  ||  |     ||  ||  | |   ||
Db       4668 GCAGTTCCTCGCCCAGGCGCAGACGCACCTCGACGCGTCCCTCACCCTCGACGTCATCGC 4727

Qy        320 GGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTC 379
               | || || | | |    || |  ||  | | ||    ||  | | ||| | |   ||
Db       4728 CGACTGCGAAACGCTGCAGGACGTCGTCGACGCGATCACCGGCACCGCCTCTGACACCGG 4787

Qy        380 CGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGG 439
              | ||||| || ||   || ||  ||      | |||||| |     ||  | |  ||||
Db       4788 CACGGCCGCCCCAAAGCCCGCTTCG------GTGGCTCCCGTTGAGGAGGCCCCGGCGGC 4841

Qy        440 GGACGCACGGCCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGA 499
              | ||||| | |   |||          | |   | |||   ||  |  | ||  |
Db       4842 CGCCGCACCGGCAAAGGCCCCGGTACGGCGCGCCGCGGCCGCATCGCCGAACGACTTCCC 4901

Qy        500 CGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGC 559
              ||| || |||  ||| |||      |  |||  || ||  |  |||||||| |||  ||
Db       4902 CGAACTGGTCCGCATGAACGGTGTCACCTCCGGCCGGCCCGTGTTCTGGGTCCACCACGG 4961

Qy        560 GCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTT 619
                ||   || | || |   |||  ||| |   ||||| | | || ||  |   || ||
Db       4962 CAACGGCGGCGTGGAGTC---CTACGCCCCGCTGGCCGCACGCTGCCCGCGT---CCCTT 5015

Qy        620 CTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGA 679
              || || | ||||| |   ||||      ||     | ||      | ||||  |      | |
Db       5016 CTACGGCATCCAGCCGAAGGGCTGGATCGACTCCACCGACATCCTCACCGGTCAGTACGC 5075

Qy        680 GATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGT 739
               || || |  || ||||   ||  | | |            |||  ||| || |||| | |
Db       5076 CATGGCCGAGCACTACGCGTCCCTCATCCTCGCCGTACAGCCGGAAGGCCCGTACGACAT 5135

Qy        740 GGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCA 799
              || ||| | |||   |||||| |  |||| | |||||  |  ||||||||||    |
Db       5136 CGGCGGGTTCTCCCTGGGCGGCCTGTTCGCGTACGAGACCGTGCGGCAGCTCCAGCTGAC 5195

Qy        800 AGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACAC 835
               || |    |||   |  ||||||||||| |||||
Db       5196 GGGCGCCGACGTGCGCACGCTGGTCATGCTGGACAC 5231


RESULT 11
AAD54217
ID   AAD54217 standard; DNA; 52101 BP.
XX
AC   AAD54217;
XX
DT   17-JUN-2003  (first entry)
XX
DE   Streptomyces platensis subspecies rosaceus dorrigocin DNA.
XX
KW   Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
KW   ds.
XX
OS   Streptomyces platensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(67. .3720)
FT                   /*tag=  a
FT                   /product= "ORF1 protein"
FT   CDS             4092. .5681
FT                   /*tag=  b
FT                   /product= "ORF2 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             5767. .6018
FT                   /*tag=  c
FT                   /product= "ORF3 protein"
FT   CDS             6023. .7993
FT                   /*tag=  d
FT                   /product= "ORF4 protein"
FT   CDS             8009. .17587
FT                   /*tag=  e
FT                   /product= "ORF5 protein"
FT   CDS             17634. .41714
FT                   /*tag=  f
FT                   /product= "ORF6 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             41772. .47633
FT                   /*tag=  g
FT                   /product= "ORF7 protein"
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             47635. .49890
FT                   /*tag=  h
FT                   /product= "ORF8 protein"
FT   CDS             49922. .50938
FT                   /*tag=  i
FT                   /product= "ORF9 protein"
FT   CDS             51234. .52079
FT                   /*tag=  j
FT                   /product= "ORF10 protein"
XX
PN   WO200288176-A2.
XX
PD   07-NOV-2002.
XX
PF   26-APR-2002; 2002WO-CA000591.
XX
PR   26-APR-2001; 2001US-0286346P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazopoulos E,  Staffa A,  Yang X;
XX
```

```
DR   WPI; 2003-201222/19.
DR   P-PSDB; AAE35484, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,
DR   AAE35490, AAE35491, AAE35492, AAE35493.
XX
PT   Novel isolated or purified polypeptide involved in biosynthesis of
PT   polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
PT   dorrigocin or lactimidomycin.
XX
PS   Claim 1; Page 85-113; 312pp; English.
XX
CC   The invention relates to novel proteins involved in the biosynthesis of
CC   polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC   microorganisms. Sequences of the invention allow direct manipulation of
CC   dorrigocin, lactimidomycin and related chemical structures via chemical
CC   engineering of the enzymes involved in the biosynthesis of dorrigocin and
CC   lactimidomycin. They are useful for introducing chemical handles into
CC   normally inert positions that permit subsequence chemical modifications
CC   and facilitate the development of polyketides. The genes and proteins of
CC   the invention can also be used to generate a focused library of analogues
CC   around a polyketide lead candidate to fine-tune the compound for optimal
CC   properties. They are useful for generating antibodies specific for the
CC   polyketide biosynthesis. The present sequence is Streptomyces platensis
CC   subspecies rosaceus DORR DNA
XX
SQ   Sequence 52101 BP; 7364 A; 20113 C; 17894 G; 6730 T; 0 U; 0 Other;

  Query Match             10.3%;  Score 102.8;  DB 10;  Length 52101;
  Best Local Similarity   52.4%;  Pred. No. 3.3e-06;
  Matches  302;  Conservative    0;  Mismatches  262;  Indels   12;  Gaps    3;

Qy        260 GCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGC 319
              || ||   |||||  ||| | | ||  |  ||  ||  |      ||  ||  | |   ||
Db      46439 GCAGTTCCTCGCCCAGGCGCAGACGCACCTCGACGCGTCCCTCACCCTCGACGTCATCGC 46498

Qy        320 GGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTC 379
              | || || | | |    || |  ||  | | ||    ||  | | ||| | |   ||
Db      46499 CGACTGCGAAACGCTGCAGGACGTCGTCGACGCGATCACCGGCACCGCCTCTGACACCGG 46558

Qy        380 CGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGG 439
              | ||||| || ||   || ||  ||      | |||||| |     ||  | |  ||||
Db      46559 CACGGCCGCCCCAAAGCCCGCTTCG------GTGGCTCCCGTTGAGGAGGCCCCGGCGGC 46612

Qy        440 GGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGA 499
              | |||||  | |   |||           | |   | |||   ||  |  | ||  |
Db      46613 CGCCGCACCGGCAAAGGCCCCGGTACGGCGCGCCGCGGCCGCATCGCCGAACGACTTCCC 46672

Qy        500 CGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGC 559
              ||| || |||  ||| |||      |  |||  || ||  |  |||||||| |||  ||
Db      46673 CGAACTGGTCCGCATGAACGGTGTCACCTCCGGCCGGCCCGTGTTCTGGGTCCACCACGG 46732

Qy        560 GCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTT 619
                ||   || | || |   |||  ||| |   ||||| | | || ||  |   || ||
Db      46733 CAACGGCGGCGTGGAGTC---CTACGCCCCGCTGGCCGCACGCTGCCCGCGT---CCCTT 46786

Qy        620 CTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGA 679
              || || | ||||| |   ||||      ||    | ||      | ||||  |      | |
Db      46787 CTACGGCATCCAGCCGAAGGGCTGGATCGACTCCACCGACATCCTCACCGGTCAGTACGC 46846

Qy        680 GATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGT 739
              || || |  || ||||   ||  | | |           |||  ||| || |||| | |
Db      46847 CATGGCCGAGCACTACGCGTCCCTCATCCTCGCCGTACAGCCGGAAGGCCCGTACGACAT 46906

Qy        740 GGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCA 799
              || ||| | |||   |||||| |  |||| | |||||  |  ||||||||||     |
Db      46907 CGGCGGGTTCTCCCTGGGCGGCCTGTTCGCGTACGAGACCGTGCGGCAGCTCCAGCTGAC 46966

Qy        800 AGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACAC 835
              || |    |||   | ||||||||||| |||||
Db      46967 GGGCGCCGACGTGCGCACGCTGGTCATGCTGGACAC 47002


RESULT 12
ABD02736/c
ID   ABD02736 standard; DNA; 1209 BP.
XX
AC   ABD02736;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #1340.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO69165.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 1340; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 1209 BP; 171 A; 389 C; 459 G; 190 T; 0 U; 0 Other;

  Query Match             10.2%;  Score 101.6;  DB 11;  Length 1209;
  Best Local Similarity   50.3%;  Pred. No. 1e-05;
  Matches  305;  Conservative    0;  Mismatches  294;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||           ||   ||
Db       1125 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 1066

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| |||||  |  |  |  |  || | ||  |  |   ||| |||     |||  |
Db       1065 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 1006

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
              | ||      ||   |  | |  ||||     |||    |||  | | |  | |  ||
Db       1005 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 950

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
              | | |  |  ||  | |    |  | | | | | | ||| || |  |
Db        949 GACGCCGACCGTGGAGGGCCTCGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 890
```

```
Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              ||||  ||| ||    |||    |  |||||   |||  |     ||||   || |||
Db        889 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 830

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
                || | ||  |       |   || |||     ||| ||     ||||  | ||
Db        829 GCTCGGCGGCCACGTACTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 770

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||   |||  || |  |  |   | |||  |  ||  | ||||
Db        769 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 710

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              ||| ||||| |||  |||| ||||     | || | |       |||  ||| || ||
Db        709 GGACATCGCCGCCAGTTACCTCGCGGCGATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 650

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| ||||||||||||||| || | | |
Db        649 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 590

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |   | ||   ||| |||  ||| || |||||||| ||    |||  | |
Db        589 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 530

Qy        854 CCTCGC 859
              || |||
Db        529 CCACGC 524
```

RESULT 13
ABD02659
ID ABD02659 standard; DNA; 1260 BP.
XX
AC ABD02659;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #1263.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO69088.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 1263; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1260 BP; 199 A; 475 C; 407 G; 179 T; 0 U; 0 Other;

 Query Match 10.2%; Score 101.6; DB 11; Length 1260;
 Best Local Similarity 50.3%; Pred. No. 9.9e-06;
 Matches 305; Conservative 0; Mismatches 294; Indels 7; Gaps 2;

```
Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||             ||   ||
Db        157 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 216

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| |||||  |  |  |  |  || | ||  |  |   ||| |||      |||  |
Db        217 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 276

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
              | ||      ||   |  | |   ||||     |||    |||  | | |  | |  ||
Db        277 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 332

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
              | |    |   |   ||    | |      |   | | | | |  | ||| || |  |
Db        333 GACGCCGACCGTGGAGGGCCTCGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 392

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              ||||  ||| ||    |||    |  |||||   |||  |     ||||   || |||
Db        393 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 452

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
                || | ||  |       |   || |||     ||| ||     ||||  | ||
Db        453 GCTCGGCGGCCACGTACTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 512

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |   | |||  |  ||  | ||||
Db        513 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 572

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              ||| ||||| |||  |||| ||||     | || | |      |||  ||| || ||
Db        573 GGACATCGCCGCCAGTTACCTCGCGGCGATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 632

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| ||||||||||||||| || | | |
Db        633 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 692

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |   | ||   ||| |||  ||| || |||||||| ||    |||  | |
Db        693 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 752

Qy        854 CCTCGC 859
              || |||
Db        753 CCACGC 758
```

RESULT 14
ABD02615
ID ABD02615 standard; DNA; 6858 BP.
XX
AC ABD02615;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #1219.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;

```
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO69044.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 1219; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 6858 BP; 1005 A; 2580 C; 2233 G; 1040 T; 0 U; 0 Other;

  Query Match            10.2%;  Score 101.6;  DB 11;  Length 6858;
  Best Local Similarity  50.3%;  Pred. No. 7.2e-06;
  Matches  305;  Conservative    0;  Mismatches  294;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||           ||   ||
Db       6164 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6223

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| |||||  |  |  |  |  || | ||  |  |   ||| |||     |||  |
Db       6224 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6283

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
              | |       ||   |  | |   ||||     |||    |||  | | |  | |  ||
Db       6284 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6339

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
              | |     |   ||   | |     |  | | | | |  | ||| || |  |
Db       6340 GACGCCGACCGTGGAGGGCCTCGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6399

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              ||||  ||| ||    |||    |  |||||   |||  |     ||||   || |||
Db       6400 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6459

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
                    || | ||  |       |   || |||     ||| ||     ||||  | ||
Db       6460 GCTCGGCGGCCACGTACTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6519

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |    | |||  |  ||  | ||||
Db       6520 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6579

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              ||| ||||| |||  |||| ||||      | || | |       |||  ||| || ||
Db       6580 GGACATCGCCGCCAGTTACCTCGCGGCGATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6639

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| |||||||||||||| || | | |
Db       6640 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6699

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |   | ||   ||| |||  ||| || |||||||| ||    |||  | |
Db       6700 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6759

Qy        854 CCTCGC 859
              || |||
Db       6760 CCACGC 6765


RESULT 15
ACA26601
ID   ACA26601 standard; DNA; 10002 BP.
XX
AC   ACA26601;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #8258.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Burkholderia mallei.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU22731.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 14471; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
```

CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 10002 BP; 1407 A; 3398 C; 3734 G; 1463 T; 0 U; 0 Other;

Query Match 10.1%; Score 100.8; DB 8; Length 10002;
Best Local Similarity 47.7%; Pred. No. 8.6e-06;
Matches 356; Conservative 0; Mismatches 387; Indels 3; Gaps 2;

```
Qy       149 CGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAG 208
            ||||||| ||   |  |   |   ||  | ||||||| || ||   |   |   |   |
Db      7986 CGCCGCGTACGTGCGGCTGCCGGCGATGCCGCTCACGCCCAACGGCAAGCTCGAGCGCCG 8045

Qy       209 CAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGAT 268
            |         |||  ||    | |  ||   || ||  | || |    |  |||||  |
Db      8046 CGCGCTGCCCGCCCCCGACGAGCGATCGTACGTGCGGCGCGACTACGCGGCGCCGCAGGG 8105

Qy       269 CGCCACGGCCCCGGCGGCCTC-CGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCG 327
            ||  |  |   || |   | | || | |   |  |  ||   ||||  ||   || || |
Db      8106 CGAGATCGAGACGACACTCGCGCGGATCTGGGCGGAGCTGTTCGGCATCGAGCGCGTCGG 8165

Qy       328 CAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCC--TCCGCGGCCCCCGCGGCCTCCGCGGC 385
            | | | |  |||| ||  |   | ||||||  |  ||   |  | | | |       |  ||
Db      8166 CCGGCACGACGGCTTCTTCGAACTCGGCGGGCATTCGCTGCTCGCGGTGCGGATGGTCGC 8225

Qy       386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGC 445
               ||    |  | |  |  |  |||||  |||  ||  |  | |  |||  |     |
Db      8226 GCGCGTGCACGATGTGCTGGGCGTCGAGGTGCCGCTGCGCGCGCTGTTCGCCGATCCGGT 8285

Qy       446 ACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCT 505
             | || || |   |          | | |||   |   | || | |       |     | ||
Db      8286 GCTGCACGTGTTCGCGTCGGCGGTCGCGCGCGCGTCGACGCGCCAGGCGTCGTCGAATCT 8345

Qy       506 CGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGG 565
            |||||   ||  |    ||| |   || |     ||||   | ||   | ||   ||
Db      8346 CGTCGCGTTCCGCAGCGCGGGCACGGCCGCGCCGCTCTTCTTCATTCATTCGGGGCTCGG 8405

Qy       566 GCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGC 625
                     ||     |||    |||   |||   ||||    ||| | | ||
Db      8406 CGAGATCGGCTTCGTCGGCGATCTGCTGCCCGGCATCGCGCCGGAGATTCCGGTGTACGG 8465

Qy       626 CTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGC 685
            ||||  |||    ||  ||  || |  ||  | |||   | |||  ||||||||||| ||
Db      8466 CTTCGCGGCGGTCGGTTTCCTCGCGGGCGAGACGCCGCACGCGACGATCGAGGAGATGGC 8525

Qy       686 GGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGG 745
             || || || |||| |    ||||||       |||   ||||| ||     | |  ||
Db      8526 CGCGCAATATGTCGACGCGATGCGGCGCGTGCAGCCGCATGGGCCGTATCGGCTCGCCGG 8585

Qy       746 GCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGA 805
            |  || | | ||||||| |||||| ||||| |||||||  |||||  || ||   |  ||
Db      8586 GTGGTGCGCGGGCGGCAACATCGCGTTCGAAATGGCCCATCAGCTGATCGCGGCCGACGA 8645

Qy       806 GCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGA 865
            |   |||      ||   |||| ||||  |     |  | |   ||  |  |     ||
```

```
Db      8646 GACGGTCGAGTTCCTCTGCATGATCGATTCGCCGACATCCGCGCCGATCGACCGCTCGGT 8705

Qy       866 GACGCCGGGCGACATCGACCCGATCC 891
                ||  || |||   |||  |  || |
Db      8706 CACCGCGTGCGTGCTCGCGCGCATTC 8731
```

Search completed: April 27, 2006, 08:00:09
Job time : 487.212 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 206.315 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_55000_55999
Perfect score: 1000
Sequence: 1 tcacggcgatggtctgcgat.....cccccgatgcagctcccgcg 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		120.2	12.0	4296	3	US-09-902-540-3291	Sequence 3291, Ap
2	c	120.2	12.0	17727	3	US-09-902-540-1152	Sequence 1152, Ap
3		106.4	10.6	7110	3	US-09-653-730-26	Sequence 26, Appl
4	c	101.6	10.2	1209	3	US-09-252-991A-1340	Sequence 1340, Ap
5		101.6	10.2	1260	3	US-09-252-991A-1263	Sequence 1263, Ap
6		101.6	10.2	6858	3	US-09-252-991A-1219	Sequence 1219, Ap
7		98.6	9.9	24459	3	US-09-902-540-5004	Sequence 5004, Ap
8	c	98.6	9.9	32241	3	US-09-902-540-1247	Sequence 1247, Ap
9		98	9.8	1002	3	US-09-902-540-3539	Sequence 3539, Ap
10		98	9.8	1008	3	US-09-902-540-117	Sequence 117, App
11		97.2	9.7	2196	3	US-09-902-540-6536	Sequence 6536, Ap
12		97.2	9.7	2200	3	US-09-902-540-478	Sequence 478, App
13	c	97	9.7	462	3	US-09-252-991A-1299	Sequence 1299, Ap
14		96	9.6	3957	3	US-10-237-551-193	Sequence 193, App
15		96	9.6	154746	3	US-09-827-688-8	Sequence 8, Appli
16	c	96	9.6	154746	3	US-09-827-688-8	Sequence 8, Appli
17		95.6	9.6	2937	3	US-09-902-540-6675	Sequence 6675, Ap
18		95.6	9.6	3550	3	US-09-902-540-524	Sequence 524, App
19		92.4	9.2	3978	3	US-09-266-965-19	Sequence 19, Appl
20		92.4	9.2	12249	3	US-09-266-965-74	Sequence 74, Appl
21		92.4	9.2	18331	3	US-09-266-965-96	Sequence 96, Appl
22		90.4	9.0	6336	3	US-09-902-540-8124	Sequence 8124, Ap
23		90.4	9.0	6338	3	US-09-902-540-831	Sequence 831, App
24		89	8.9	1614	3	US-09-616-289-45	Sequence 45, Appl
25		89	8.9	1614	3	US-09-976-740-45	Sequence 45, Appl
26		89	8.9	12425	3	US-09-616-289-50	Sequence 50, Appl
27		89	8.9	12425	3	US-09-976-740-50	Sequence 50, Appl
28		88.2	8.8	2561	3	US-09-616-289-48	Sequence 48, Appl
29		88.2	8.8	2561	3	US-09-976-740-48	Sequence 48, Appl
30		87	8.7	3270	3	US-09-902-540-7024	Sequence 7024, Ap
31	c	87	8.7	4018	3	US-09-902-540-620	Sequence 620, App
32		86	8.6	1293	3	US-09-902-540-6763	Sequence 6763, Ap
33	c	86	8.6	5304	3	US-09-902-540-551	Sequence 551, App
34		85.6	8.6	1926	3	US-09-249-585A-4	Sequence 4, Appli
35		85.6	8.6	1931	2	US-09-130-114-2	Sequence 2, Appli
36		82	8.2	2481	3	US-09-894-998A-35	Sequence 35, Appl
37		82	8.2	2481	3	US-10-237-551-35	Sequence 35, Appl
38		82	8.2	3066	3	US-10-237-551-152	Sequence 152, App
39	c	81.8	8.2	1028	3	US-08-118-200-1	Sequence 1, Appli
40	c	81.8	8.2	1028	3	US-08-458-745-1	Sequence 1, Appli
41	c	80.8	8.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
42		80.4	8.0	1327	3	US-08-483-533-36	Sequence 36, Appl
43		80.4	8.0	1327	3	US-09-283-471A-36	Sequence 36, Appl
44		79.6	8.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
45	c	79.2	7.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-902-540-3291
; Sequence 3291, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3291
; LENGTH: 4296
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3291

Query Match 12.0%; Score 120.2; DB 3; Length 4296;
Best Local Similarity 52.2%; Pred. No. 6.6e-12;
Matches 314; Conservative 0; Mismatches 283; Indels 4; Gaps 2;

```
Qy       326 CGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGC 385
            ||| |  | ||   |  ||| | ||  || |     |  ||| ||| |   | | |
Db      3605 CGCTGGGCGCGGACCTGACGCTGGCCACGCTGTTCGAGGCTCCCACGCTGGAGCAGTGCG 3664

Qy       386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTC-CGGCGCCTGATCCGCGCGCGGGGGACG 444
             | ||| |  ||  |  | ||  ||  | |    |||| ||  || |  |||| ||||
Db      3665 CCGCGCTGGTGCGTGAGGCCGCCGGGATTCCCTTCACGCCGGACGCGGGTCCGGGTGACG 3724

Qy       445 CACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGC 504
             |   ||| |  ||| |   |  | |    || | | ||||  |   | |          ||
Db      3725 CGGCGCCGGCTGCGCAGGCCCCGGGCGTCAAGGCGGTGCCGAAG---AGCTGGTCTCCGC 3781

Qy       505 TCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTG 564
             | |||  |||| |   | |||||     |    ||  ||||| ||||||||||||    |
Db      3782 TGGTCACCATCCAGAAGGGCGGCAATACCCTCCCCTTCTTCTGCGTGCACGGCGCGGGCG 3841

Qy       565 GGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCG 624
             |   ||  |      ||      ||| |     ||||||||| |||  | || |||| ||
Db      3842 GCAACGTCCTCAACTTCCGCGAGCTGGCGCAGACGCTCGGCAAGGAGCAGCCCTTCTACG 3901

Qy       625 CCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCG 684
```

```
Db       3902 GGCTCCAGGCGCGTGGCGTGGACGGCAAGCTGCCTCCGGCCGACAGCATCGAGGAGATGG 3961

Qy        685 CGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTG 744

Db       3962 CGTCCATCTACCTGGAGAGCATCCGTCAGGTGCAGCCCCACGGGCCCTACCTGCTCGGCG 4021

Qy        745 GGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGG 804

Db       4022 GCTACTCCGGCGGCGGCGTGGTCGCCTACAAAATAGCGCAGCGGCTGCACGCGCTGGGCG 4081

Qy        805 AGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGG 864

Db       4082 AGCCGGTGGCGCTGGTCGCGTTCCTGGACACCTTCCACCCGTCCACCCAGGAGCGACGCC 4141

Qy        865 AGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCA 924

Db       4142 AGTCCCTCCATGAGCGCCTGCGGGAGCTGCGCCGGGAGGGCGCGGGCAGCTACGTCAGCC 4201

Qy        925 G 925

Db       4202 G 4202
```

RESULT 2
US-09-902-540-1152/c
; Sequence 1152, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1152
; LENGTH: 17727
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(17727)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1152

Query Match 12.0%; Score 120.2; DB 3; Length 17727;
Best Local Similarity 52.2%; Pred. No. 6.8e-12;
Matches 314; Conservative 0; Mismatches 283; Indels 4; Gaps 2;

```
Qy        326 CGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGC 385

Db      14123 CGCTGGGCGCGGACCTGACGCTGGCCACGCTGTTCGAGGCTCCCACGCTGGAGCAGTGCG 14064

Qy        386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTC-CGGCGCCTGATCCGCGCGCGGGGGACG 444

Db      14063 CCGCGCTGGTGCGTGAGGCCGCCGGGATTCCCTTCACGCCGGACGCGGGTCCGGGTGACG 14004

Qy        445 CACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGC 504

Db      14003 CGGCGCCGGCTGCGCAGGCCCCGGGCGTCAAGGCGGTGCCGAAG---AGCTGGTCTCCGC 13947

Qy        505 TCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTG 564

Db      13946 TGGTCACCATCCAGAAGGGCGGCAATACCCTCCCCTTCTTCTGCGTGCACGGCGCGGGCG 13887

Qy        565 GGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCG 624

Db      13886 GCAACGTCCTCAACTTCCGCGAGCTGGCGCAGACGCTCGGCAAGGAGCAGCCCTTCTACG 13827
```

```
Qy        625 CCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCG 684

Db      13826 GGCTCCAGGCGCGTGGCGTGGACGGCAAGCTGCCTCCGGCCGACAGCATCGAGGAGATGG 13767

Qy        685 CGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTG 744

Db      13766 CGTCCATCTACCTGGAGAGCATCCGTCAGGTGCAGCCCCACGGGCCCTACCTGCTCGGCG 13707

Qy        745 GGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGG 804

Db      13706 GCTACTCCGGCGGCGGCGTGGTCGCCTACAAAATAGCGCAGCGGCTGCACGCGCTGGGCG 13647

Qy        805 AGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGG 864

Db      13646 AGCCGGTGGCGCTGGTCGCGTTCCTGGACACCTTCCACCCGTCCACCCAGGAGCGACGCC 13587

Qy        865 AGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCA 924

Db      13586 AGTCCCTCCATGAGCGCCTGCGGGAGCTGCGCCGGGAGGGCGCGGGCAGCTACGTCAGCC 13527

Qy        925 G 925

Db      13526 G 13526
```

RESULT 3
US-09-653-730-26
; Sequence 26, Application US/09653730
; Patent No. 6855513
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Lee, Kimberly
; APPLICANT: Greenberg, E. P.
; APPLICANT: Muh, Ute
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038
; CURRENT APPLICATION NUMBER: US/09/653,730
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153,022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 7110
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-653-730-26

Query Match 10.6%; Score 106.4; DB 3; Length 7110;
Best Local Similarity 50.8%; Pred. No. 1.4e-09;
Matches 308; Conservative 0; Mismatches 291; Indels 7; Gaps 2;

```
Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316

Db       6007 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6066

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376

Db       6067 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6126

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436

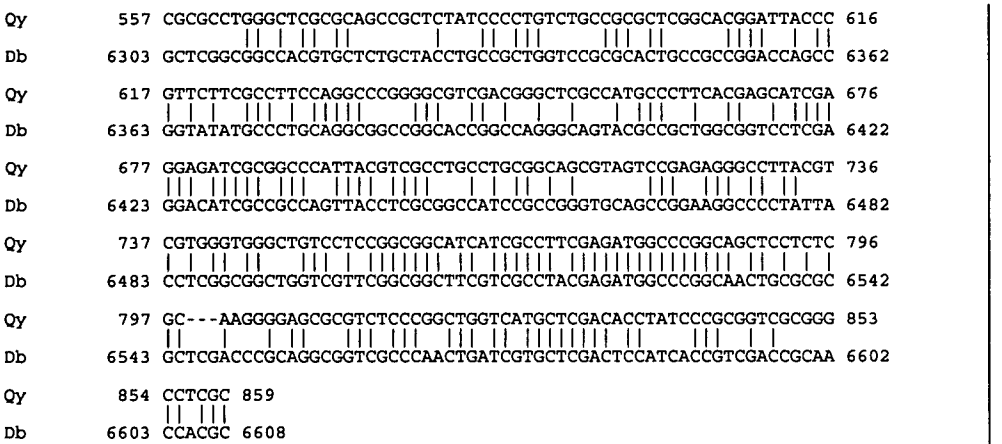
Db       6127 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6182

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496

Db       6183 GACGCCGACCGTGGAGGGCCTGGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6242

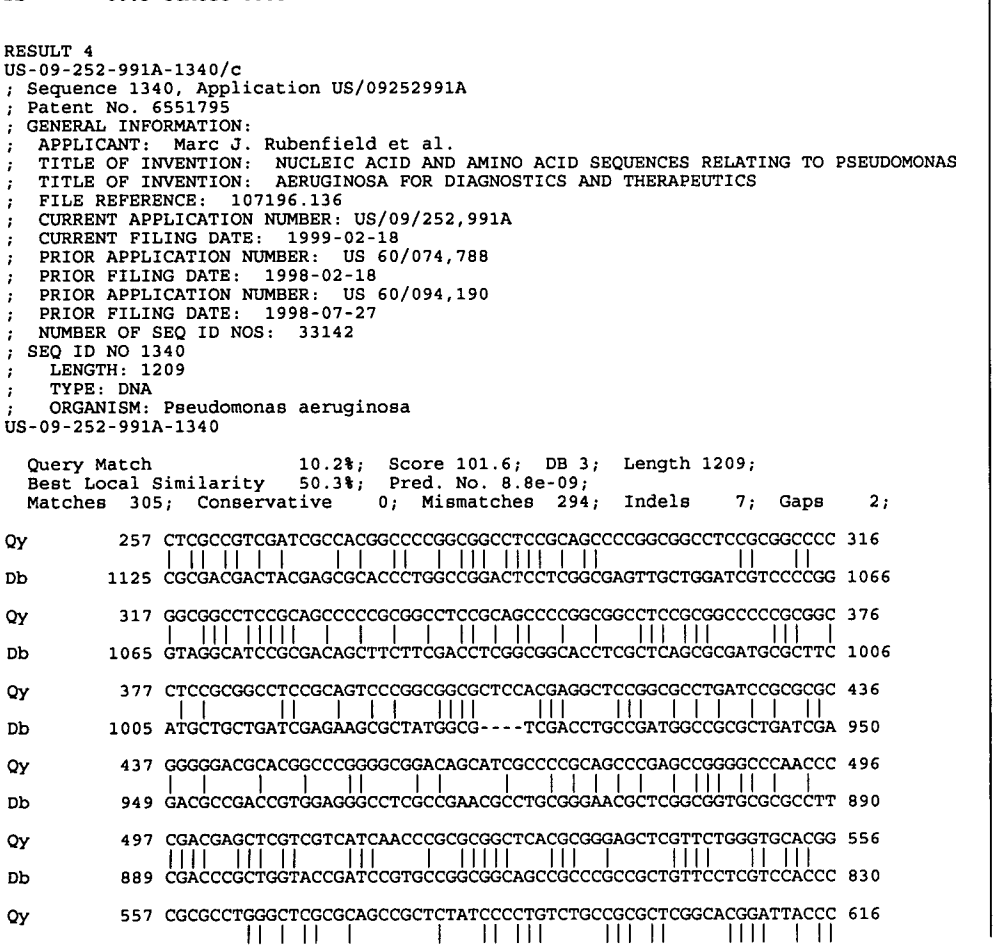
Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556

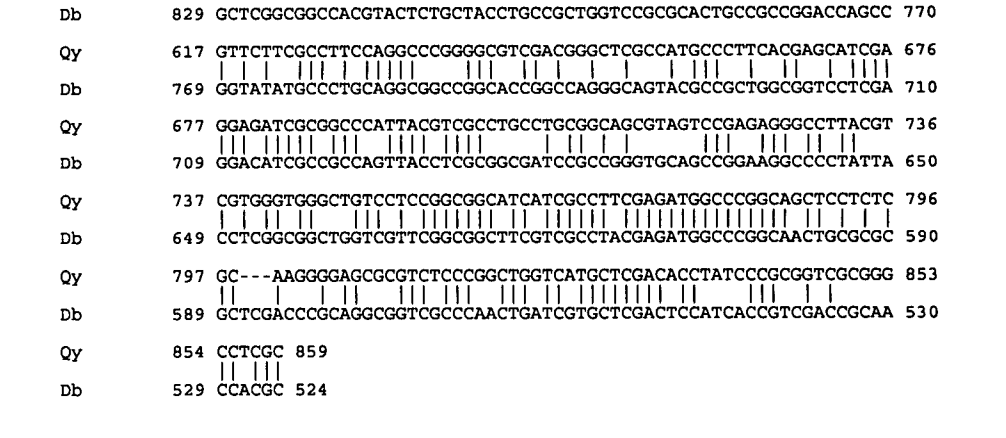
Db       6243 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6302
```



RESULT 4
US-09-252-991A-1340/c
; Sequence 1340, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1340
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1340

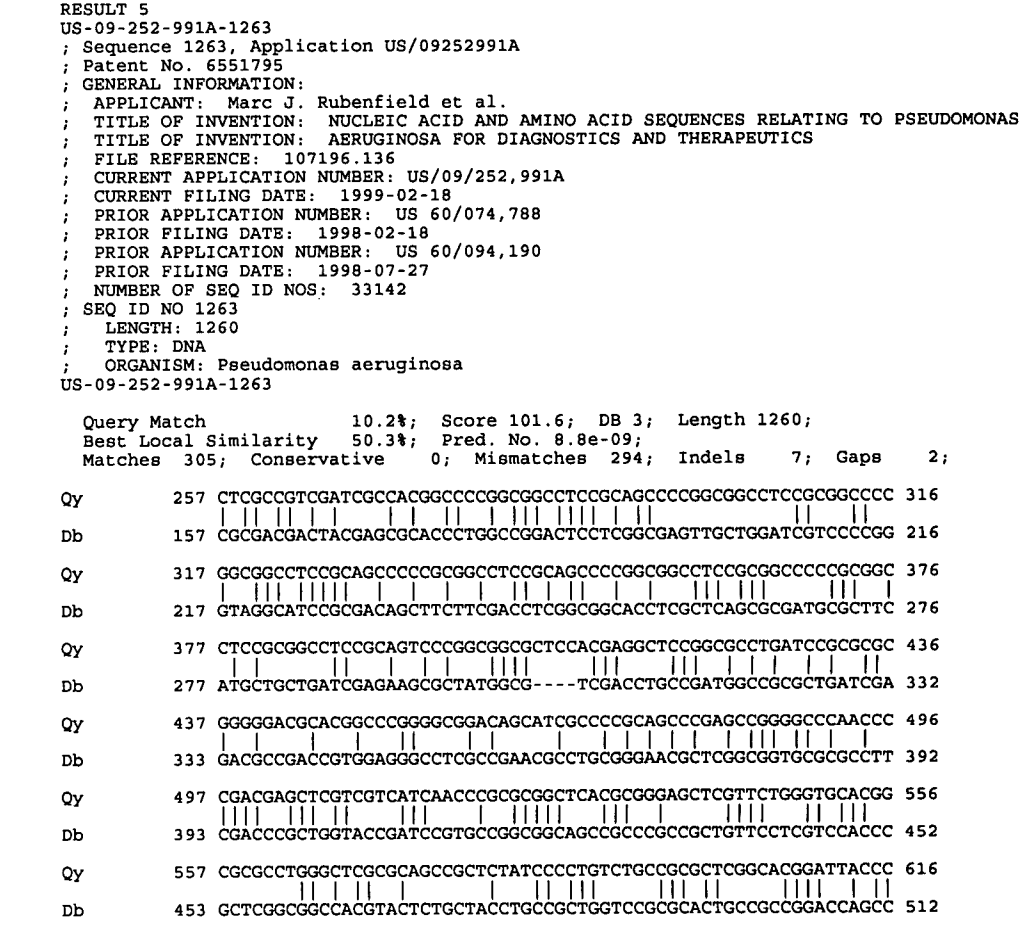
Query Match 10.2%; Score 101.6; DB 3; Length 1209;
Best Local Similarity 50.3%; Pred. No. 8.8e-09;
Matches 305; Conservative 0; Mismatches 294; Indels 7; Gaps 2;





RESULT 5
US-09-252-991A-1263
; Sequence 1263, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1263
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1263

Query Match 10.2%; Score 101.6; DB 3; Length 1260;
Best Local Similarity 50.3%; Pred. No. 8.8e-09;
Matches 305; Conservative 0; Mismatches 294; Indels 7; Gaps 2;



```
Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |    | |||  |  ||  | ||||
Db        513 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 572

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              |||  |||||  |||  ||||  ||||      | ||  | |       |||  ||| || ||
Db        573 GGACATCGCCGCCAGTTACCTCGCGGCGATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 632

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | |  ||  ||    |||  | |||||||  ||  ||||||  ||||||||||||||||  ||  | | |
Db        633 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 692

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||    |    | ||   ||| |||   ||| ||  |||||||| ||    |||   | |
Db        693 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 752

Qy        854 CCTCGC 859
              || |||
Db        753 CCACGC 758


RESULT 6
US-09-252-991A-1219
; Sequence 1219, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1219
;   LENGTH: 6858
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1219

  Query Match            10.2%;  Score 101.6;  DB 3;  Length 6858;
  Best Local Similarity   50.3%;  Pred. No. 9.1e-09;
  Matches  305;  Conservative    0;  Mismatches  294;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||           ||   ||
Db       6164 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6223

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| |||||  |  |  |  | || | ||  |  |  ||| |||     |||  |
Db       6224 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6283

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
              | |       ||   |  | |   ||||     |||    |||  | | |  | |  ||
Db       6284 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6339

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
              |  |    |   |   ||    | |     |  | | | | |  | ||| || |  |
Db       6340 GACGCCGACCGTGGAGGGCCTCGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6399

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              ||||  ||| ||    |||    |  |||||   |||  |     ||||   || |||
Db       6400 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6459

Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
              |||  | ||  |       |   || |||     ||| ||     ||||  | ||
Db       6460 GCTCGGCGGCCACGTACTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6519

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |    | |||  |  ||  | ||||
```

```
Db       6520 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6579

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              |||  |||||  |||  ||||  ||||      | ||  | |       |||  ||| || ||
Db       6580 GGACATCGCCGCCAGTTACCTCGCGGCGATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6639

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | |  ||  ||    |||  | |||||||  ||  ||||||  ||||||||||||||||  ||  | | |
Db       6640 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6699

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||    |    | ||   ||| |||   ||| ||  |||||||| ||    |||   | |
Db       6700 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6759

Qy        854 CCTCGC 859
              || |||
Db       6760 CCACGC 6765


RESULT 7
US-09-902-540-5004
; Sequence 5004, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5004
;   LENGTH: 24459
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-5004

  Query Match             9.9%;  Score 98.6;  DB 3;  Length 24459;
  Best Local Similarity   52.9%;  Pred. No. 3e-08;
  Matches  212;  Conservative    0;  Mismatches  189;  Indels    0;  Gaps    0;

Qy        442 ACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACG 501
              | ||  || |  ||||||      | | | ||  |   | |  |  ||      |
Db      23600 AGGCGGGGACGCTGGCGGAGCTGGCCTCTCGGCTCCGGCACGCACCCGGACACGTGTCCT 23659

Qy        502 AGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGC 561
              |||  ||   |||| |     | || || |   ||  |   |||||  || |||   |
Db      23660 CGCTGGTGCCCATCCAGAGTGGTGGTTCGCAGCGGCCCGTCTTCTGCGTCCACCCTGTCA 23719

Qy        562 CTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCT 621
              ||   ||  | ||||   |      |||||       || ||  ||||  | |||||||
Db      23720 GCGGCAACGTCCTGCCGTATTTGGAACTGTCGCGGCGCCTGGGGCCGGACCAACCGTTCT 23779

Qy        622 TCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGA 681
              ||| ||||||||||  |||| |||| ||         ||  |   ||  | | |||| ||
Db      23780 ACGCGTTCCAGGCCCCGGGCCTCGAAGGCGAGCGGGCTCCGCTGGGGACCGTGGAGGCGA 23839

Qy        682 TCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGG 741
              | ||  ||   || || |     | ||||           | || ||||| |||    |||
Db      23840 TGGCCTCCACCTATGTGGAGGCCATGCGCGGCGTCCAGGCCAGCGGGCCATACCGGCTGG 23899

Qy        742 GTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAG 801
              |||  |||||  || ||  |  | ||||||| ||||| ||||||  |       ||| |
Db      23900 CGGGGTGGTCCTTGGGAGGTGTGGTGGCCTTCGAAATGGCGCGGCAGTTGGAGCAGCAGG 23959

Qy        802 GGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCC 842
              ||||||  ||    | ||||    |  ||||| ||||  ||
Db      23960 GGGAGCAGGTGGAGCAGCTGACGCTCATCGACGCCTACGCC 24000
```

```
RESULT 8
US-09-902-540-1247/c
; Sequence 1247, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1247
;   LENGTH: 32241
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(32241)
;   OTHER INFORMATION: unsure at all n locations
US-09-902-540-1247

  Query Match             9.9%;  Score 98.6;  DB 3;  Length 32241;
  Best Local Similarity   52.9%;  Pred. No. 3e-08;
  Matches  212;  Conservative    0;  Mismatches  189;  Indels    0;  Gaps    0;

Qy        442 ACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACG 501
              | ||  || |   ||||||       | | | ||  |   | |          ||     |
Db       8642 AGGCGGGGACGCTGGCGGAGCTGGCCTCTCGGCTCCGGCACGCACCCGGACACGTGTCCT 8583

Qy        502 AGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGC 561
              ||| ||   |||| |     | || || |   ||  |   ||||| || |||   |
Db       8582 CGCTGGTGCCCATCCAGAGTGGTGGTTCGCAGCGGCCCGTCTTCTGCGTCCACCCTGTCA 8523

Qy        562 CTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCT 621
              ||   ||  | ||||   |     |||||       || ||  ||||  | ||||||||
Db       8522 GCGGCAACGTCCTGCCGTATTTGGAACTGTCGCGGCGCCTGGGGCCGGACCAACCGTTCT 8463

Qy        622 TCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGA 681
              ||| ||||||||||| |||| |||| ||          ||  |   || | | |||| ||
Db       8462 ACGCGTTCCAGGCCCCGGGCCTCGAAGGCGAGCGGGCTCCGCTGGGGACCGTGGAGGCGA 8403

Qy        682 TCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGG 741
              | ||  ||   || || |    | ||||          | || ||||| |||    |||
Db       8402 TGGCCTCCACCTATGTGGAGGCCATGCGCGGCGTCCAGGCCAGCGGGCCATACCGGCTGG 8343

Qy        742 GTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAG 801
              |||  |||||  || ||  |  | ||||||| ||||| ||||||  |      ||| |
Db       8342 CGGGGTGGTCCTTGGGAGGTGTGGTGGCCTTCGAAATGGCGCGGCAGTTGGAGCAGCAGG 8283

Qy        802 GGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCC 842
              ||||||  ||    | ||||    |  ||||| ||||  ||
Db       8282 GGGAGCAGGTGGAGCAGCTGACGCTCATCGACGCCTACGCC 8242


RESULT 9
US-09-902-540-3539
; Sequence 3539, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3539
;   LENGTH: 1002
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-3539

  Query Match             9.8%;  Score 98;  DB 3;  Length 1002;
  Best Local Similarity   56.9%;  Pred. No. 3.6e-08;
  Matches  199;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy        516 AACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAG 575
              | |||  | | | ||  | |  |||  | || |   | | |  |||  |   ||  || |
Db        239 AGCCCCTGGGACACAAACCGCCGCTGTTCCTTGCCCCGCCGTCCGCTGGCAGTCCGGCCG 298

Qy        576 CCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCC 635
                |    | |||  | | ||| | || | | ||||  | ||  | |||| |||||||
Db        299 TCTACGTCACCCTGGCCCGCCAC-CTGGACGCGGAGCAGCCCGTGTTCGGCTTCCAGATG 357

Qy        636 CGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTAC 695
              | ||||||   |    |    ||||   |  || ||||||||| | ||| || |  ||
Db        358 CCTGGCGTCATGGACGACCAACTGCCACCCGAGACCATCGAGGACACCGCCGCGCTGTAT 417

Qy        696 GTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCC 755
              || ||   |||||||||       ||| | ||||| |||    | |  ||   ||||| |
Db        418 GTGGCGGCCCTGCGGCAGGTCCAGCCGCGCGGGCCGTACCGGATCGCGGGCTGGTCCTAC 477

Qy        756 GGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCC 815
              || ||  || ||| ||  || ||||| ||||||||    |||  ||||| |  ||  |
Db        478 GGAGGACTCGTCGTCTGTGAAATGGCGCGGCAGCTCGAAGCGCTCGGGGAACAGGTGGCG 537

Qy        816 CGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGA 865
              | ||| | | || | |||  |    |   || |  |     ||||| ||
Db        538 CTGCTCGGCTTGATTGACGGCGCGGCGTTGGACCGGTTGTCCGCGCACGA 587


RESULT 10
US-09-902-540-117
; Sequence 117, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 117
;   LENGTH: 1008
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-117

  Query Match             9.8%;  Score 98;  DB 3;  Length 1008;
  Best Local Similarity   56.9%;  Pred. No. 3.6e-08;
  Matches  199;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy        516 AACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAG 575
              | |||  | | | ||  | |  |||  | || |   | | |  |||  |   ||  || |
Db        240 AGCCCCTGGGACACAAACCGCCGCTGTTCCTTGCCCCGCCGTCCGCTGGCAGTCCGGCCG 299

Qy        576 CCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCC 635
                |    | |||  | | ||| | || | | ||||  | ||  | |||| |||||||
Db        300 TCTACGTCACCCTGGCCCGCCAC-CTGGACGCGGAGCAGCCCGTGTTCGGCTTCCAGATG 358

Qy        636 CGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTAC 695
```

```
Db        359 CCTGGCGTCATGGACGACCAACTGCCACCCGAGACCATCGAGGACACCGCCGCGCTGTAT 418

Qy        696 GTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCC 755

Db        419 GTGGCGGCCCTGCGGCAGGTCCAGCCGCGCGGGCCGTACCGGATCGCGGGCTGGTCCTAC 478

Qy        756 GGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCC 815

Db        479 GGAGGACTCGTCGTCTGTGAAATGGCGCGGCAGCTCGAAGCGCTCGGGGAACAGGTGGCG 538

Qy        816 CGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGA 865

Db        539 CTGCTCGGCTTGATTGACGGCGCGGCGTTGGACCGGTTGTCCGCGCACGA 588


RESULT 11
US-09-902-540-6536
; Sequence 6536, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6536
;   LENGTH: 2196
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-6536

  Query Match             9.7%;  Score 97.2;  DB 3;  Length 2196;
  Best Local Similarity   54.9%;  Pred. No. 4.9e-08;
  Matches  192;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        543 TTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTC 602

Db       1610 TTCTGCGTGCACCCTGTCGGAGGCAACGTCCTGGCCTACGCGGAGCTGGCCCGGCTGCTG 1669

Qy        603 GGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCC 662

Db       1670 GGGCCGGAGCGCCCGTTCTACGGACTCCAGGCGCAAGGGCTCGACGGCAGCACGCCGCCG 1729

Qy        663 TTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCG 722

Db       1730 CTGGGCACGGTGGAGGAGATGGCCGCCGCCTATGTGGAGGCCATCCGCACGGTCCAACCC 1789

Qy        723 AGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCC 782

Db       1790 GCCGGGCCCTACCTGCTGGGCGGCTGGTCCGTGGGCGGCGTCATTGCCTATGAAATGGCG 1849

Qy        783 CGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCC 842

Db       1850 CGCCAGCTTCGCGAGCGTGGCGACACGGTGGCCCTGTTGGCGCTCATCGACGCCTACACG 1909

Qy        843 GCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCT 892

Db       1910 CCCGCGGTGCTCAACGCCGCCGAACCGGAGCCCGACGCCGCGCAGGTCGT 1959


RESULT 12
US-09-902-540-478
; Sequence 478, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 478
;   LENGTH: 2200
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-478

  Query Match             9.7%;  Score 97.2;  DB 3;  Length 2200;
  Best Local Similarity   54.9%;  Pred. No. 4.9e-08;
  Matches  192;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        543 TTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTC 602

Db       1614 TTCTGCGTGCACCCTGTCGGAGGCAACGTCCTGGCCTACGCGGAGCTGGCCCGGCTGCTG 1673

Qy        603 GGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCC 662

Db       1674 GGGCCGGAGCGCCCGTTCTACGGACTCCAGGCGCAAGGGCTCGACGGCAGCACGCCGCCG 1733

Qy        663 TTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCG 722

Db       1734 CTGGGCACGGTGGAGGAGATGGCCGCCGCCTATGTGGAGGCCATCCGCACGGTCCAACCC 1793

Qy        723 AGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCC 782

Db       1794 GCCGGGCCCTACCTGCTGGGCGGCTGGTCCGTGGGCGGCGTCATTGCCTATGAAATGGCG 1853

Qy        783 CGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCC 842

Db       1854 CGCCAGCTTCGCGAGCGTGGCGACACGGTGGCCCTGTTGGCGCTCATCGACGCCTACACG 1913

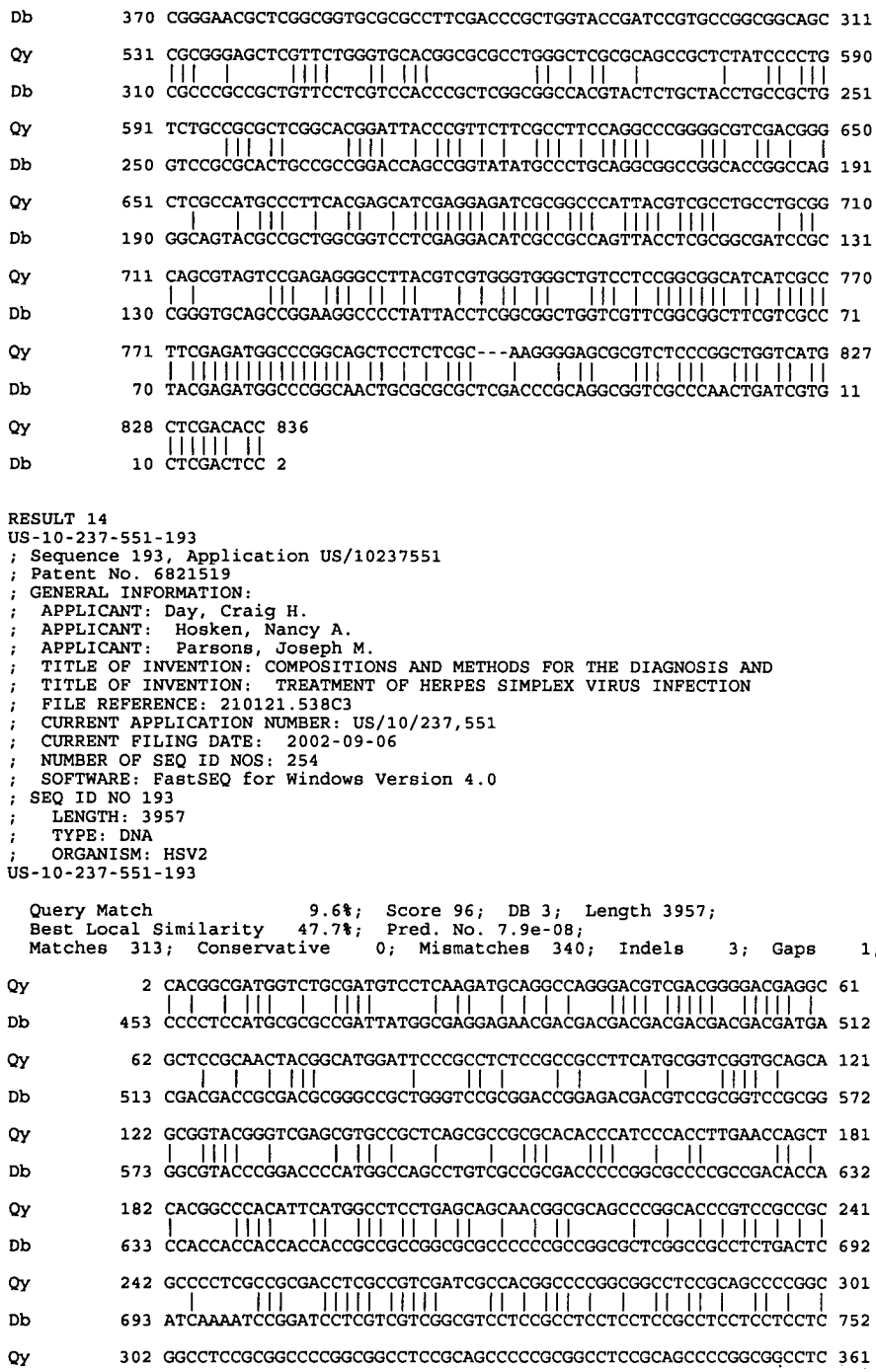
Qy        843 GCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCT 892

Db       1914 CCCGCGGTGCTCAACGCCGCCGAACCGGAGCCCGACGCCGCGCAGGTCGT 1963


RESULT 13
US-09-252-991A-1299/c
; Sequence 1299, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1299
;   LENGTH: 462
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1299

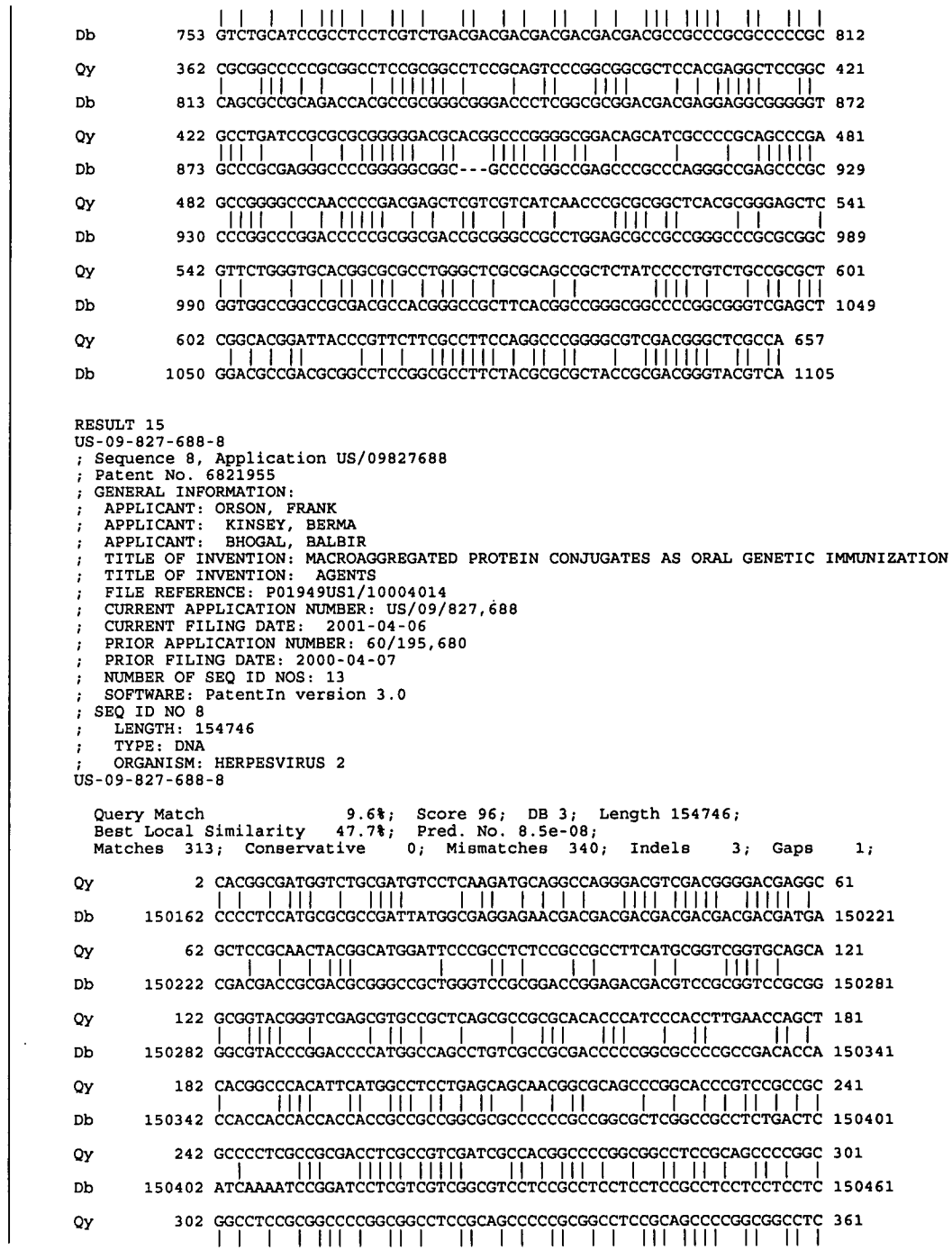
  Query Match             9.7%;  Score 97;  DB 3;  Length 462;
  Best Local Similarity   55.8%;  Pred. No. 5.2e-08;
  Matches  206;  Conservative    0;  Mismatches  160;  Indels    3;  Gaps    1;

Qy        471 CCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCA 530
```



RESULT 14
US-10-237-551-193
; Sequence 193, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 3957
; TYPE: DNA
; ORGANISM: HSV2
US-10-237-551-193

Query Match 9.6%; Score 96; DB 3; Length 3957;
Best Local Similarity 47.7%; Pred. No. 7.9e-08;
Matches 313; Conservative 0; Mismatches 340; Indels 3; Gaps 1;



RESULT 15
US-09-827-688-8
; Sequence 8, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERMA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION D.
; TITLE OF INVENTION: AGENTS
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 9.6%; Score 96; DB 3; Length 154746;
Best Local Similarity 47.7%; Pred. No. 8.5e-08;
Matches 313; Conservative 0; Mismatches 340; Indels 3; Gaps 1;

```
Db     150462 GTCTGCATCCGCCTCCTCGTCTGACGACGACGACGACGACGCCGCCCGCGCCCCCGC 150521

Qy        362 CGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGC 421

Db     150522 CAGCGCCGCAGACCACGCCGCGGGCGGGACCCTCGGCGCGGACGACGAGGAGGCGGGGGT 150581

Qy        422 GCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGA 481

Db     150582 GCCCGCGAGGGCCCCGGGGGCGGC---GCCCCGGCCGAGCCCGCCCAGGGCCGAGCCCGC 150638

Qy        482 GCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTC 541

Db     150639 CCCGGCCCGGACCCCCGCGGCGACCGCGGGCCGCCTGGAGCGCCGCCGGGCCCGCGCGGC 150698

Qy        542 GTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCT 601

Db     150699 GGTGGCCGGCCGCGACGCCACGGGCCGCTTCACGGCCGGGCGGCCCCGGCGGGTCGAGCT 150758

Qy        602 CGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCA 657

Db     150759 GGACGCCGACGCGGCCTCCGGCGCCTTCTACGCGCGCTACCGCGACGGGTACGTCA 150814
```

Search completed: April 27, 2006, 12:42:38
Job time : 209.315 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 1002.57 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_55000_55999
Perfect score: 1000
Sequence: 1 tcacggcgatggtctgcgat.....cccccgatgcagctcccgcg 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
c	2	121.4	12.1	9025608	6	US-10-156-761-1	Sequence 1, Appli
	3	112.4	11.2	1257	6	US-10-156-761-5999	Sequence 5999, Ap
	4	110	11.0	1121	7	US-10-282-122A-13806	Sequence 13806, A
	5	107.8	10.8	7140	8	US-10-211-028-3	Sequence 3, Appli
	6	107.8	10.8	90597	8	US-10-211-028-1	Sequence 1, Appli
	7	106.4	10.6	7059	7	US-10-389-647-202	Sequence 202, App
	8	106.4	10.6	7110	10	US-11-014-621-26	Sequence 26, Appl
	9	102.8	10.3	5862	6	US-10-132-134-15	Sequence 15, Appl
	10	102.8	10.3	52101	6	US-10-132-134-1	Sequence 1, Appli
c	11	102.2	10.2	972	8	US-10-425-115-17653	Sequence 17653, A
	12	101.2	10.1	810	6	US-10-156-761-3145	Sequence 3145, Ap
	13	100.8	10.1	10002	7	US-10-282-122A-14471	Sequence 14471, A
	14	98	9.8	6297	6	US-10-132-134-37	Sequence 37, Appl
	15	98	9.8	50543	6	US-10-132-134-25	Sequence 25, Appl
c	16	96.2	9.6	3024	9	US-10-939-036-35	Sequence 35, Appl
	17	96	9.6	3957	6	US-10-200-562-193	Sequence 193, App
	18	96	9.6	3957	6	US-10-237-551-193	Sequence 193, App
	19	96	9.6	3957	9	US-10-945-050-193	Sequence 193, App
	20	96	9.6	154746	3	US-09-827-688-8	Sequence 8, Appli
c	21	96	9.6	154746	3	US-09-827-688-8	Sequence 8, Appli
	22	93.2	9.3	1337	9	US-10-779-543-5579	Sequence 5579, Ap
	23	92.4	9.2	2274	7	US-10-437-963-28837	Sequence 28837, A
	24	92.4	9.2	3978	3	US-09-953-348-19	Sequence 19, Appl
	25	92.4	9.2	3978	6	US-10-267-255-19	Sequence 19, Appl
	26	92.4	9.2	12249	3	US-09-953-348-74	Sequence 74, Appl
	27	92.4	9.2	12249	6	US-10-267-255-74	Sequence 74, Appl
	28	92.4	9.2	18331	3	US-09-953-348-96	Sequence 96, Appl
	29	92.4	9.2	18331	6	US-10-267-255-96	Sequence 96, Appl
c	30	92.2	9.2	3133	6	US-10-017-161-1483	Sequence 1483, Ap
c	31	92.2	9.2	3133	6	US-10-292-798-1191	Sequence 1191, Ap
c	32	91.6	9.2	12733	5	US-10-032-393-47	Sequence 47, Appl
c	33	91.6	9.2	12739	5	US-10-032-393-8	Sequence 8, Appli
	34	91.4	9.1	88421	3	US-09-976-059-1	Sequence 1, Appli
	35	91.2	9.1	658	7	US-10-767-701-4204	Sequence 4204, Ap
	36	90.6	9.1	673	8	US-10-425-115-11145	Sequence 11145, A
	37	89.8	9.0	645	7	US-10-767-701-3807	Sequence 3807, Ap
	38	89.8	9.0	925	7	US-10-437-963-44536	Sequence 44536, A
	39	89.6	9.0	563	7	US-10-767-701-4502	Sequence 4502, Ap
	40	89.6	9.0	9025608	6	US-10-156-761-1	Sequence 1, Appli
c	41	89.4	8.9	1117	6	US-10-017-161-1403	Sequence 1403, Ap
c	42	89.4	8.9	1117	6	US-10-292-798-1141	Sequence 1141, Ap
	43	89.4	8.9	125401	7	US-10-203-295-35	Sequence 35, Appl
	44	89	8.9	1614	3	US-09-976-740-45	Sequence 45, Appl
	45	89	8.9	1614	5	US-10-023-529-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 1000; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 1e-223;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TCACGGCGATGGTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55000 TCACGGCGATGGTCTGCGATGTCCTCAAGATGCAGGCCAGGGACGTCGACGGGGACGAGG 55059

Qy         61 CGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55060 CGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGC 55119

Qy        121 AGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55120 AGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCACCTTGAACCAGC 55179
```

```
Qy        181 TCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55180 TCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCG 55239

Qy        241 CGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55240 CGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGG 55299

Qy        301 CGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55300 CGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCT 55359

Qy        361 CCGCGGCCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55360 CCGCGGCCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGG 55419

Qy        421 CGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55420 CGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCG 55479

Qy        481 AGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55480 AGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCT 55539

Qy        541 CGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55540 CGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGC 55599

Qy        601 TCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55600 TCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGC 55659

Qy        661 CCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55660 CCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTC 55719

Qy        721 CGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55720 CGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGG 55779

Qy        781 CCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55780 CCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATC 55839

Qy        841 CCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55840 CCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCC 55899

Qy        901 TGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55900 TGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGATCAAGCCCG 55959

Qy        961 ACGACCTCGCCGGGCTCCCCCCCCCCGATGCAGCTCCCGCG 1000
              ||||||||||||||||||||||||||||||||||||||||
Db      55960 ACGACCTCGCCGGGCTCCCCCCCCCCGATGCAGCTCCCGCG 55999


RESULT 2
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             12.1%;  Score 121.4;  DB 6;  Length 9025608;
  Best Local Similarity   47.2%;  Pred. No. 1.2e-19;
  Matches  434;  Conservative    0;  Mismatches  481;  Indels    4;  Gaps    2;

Qy         54 GACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCCTTCATGCGGTCG 113
              | |||   |  ||  | |  || |        |||||  |  ||  ||  |  ||||    |
Db    3934877 GGCGACCAGGCCCCGACCCTCGCCGAACTCCGCCGGCCACCTCGGCGAGTCGATGCCCGAG 3934818

Qy        114 GTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACCCATCCCCACCTTG 173
                |  || |  ||  | |     || |  |      | ||| ||| |   |||    |
Db    3934817 TACATGGTGCCGACCGCCTGGGTGAAGCTGGACGAGTTCCCGCTCACGCTCGCCAAGAAG 3934758

Qy        174 AACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCAGCCCGGCACCCG 233
              | | | |     || |        |              ||  | ||||  | |  | |  |
Db    3934757 ATCGACCGGCACGCGCTGCCCGAGCCGTCGTACGGCGCCGACGGTACGGAAGGCCGGGCC 3934698

Qy        234 TCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCC-CCGGCGGCCTCCGC 292
              | || ||| |||  | |  | |||  |  |  | |||||  ||||  ||  |    |||
Db    3934697 TACGTCGCCCCCGCCACGCCCACCGAGGAGCGGGTCGCCGAGGCCTTCGCCACGGTCCTG 3934638

Qy        293 AGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCC 352
              |   |||     || |||    || | || ||  |   | | |||||  |    ||
Db    3934637 AAGGCCGATCAGATCAGCGCGGACGCCAGCTTCTTCGAACTCGGCGGCAACTCGCTCCAG 3934578

Qy        353 GGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGA 412
              |       |  |||   |  | | | || ||  |  |  |  ||  |
Db    3934577 GCGATGCGGGCGGTCAGCCGGATCAACAAGGCCTTCGGCGTGAAGATCAACGTCCGGATG 3934518

Qy        413 GGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCC 472
              || ||||| |     ||| | |||| |     |  ||||   | || || ||    | ||
Db    3934517 CTCTACGGCGGCGCCACCGTGGGCGGCGTGGCCGAGGCCATCGACGAACTGCTCGCCGCC 3934458

Qy        473 GCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACG 532
              |||   | | | | |  || |   |  || ||    ||  |  |  |  |||| | |
Db    3934457 AAGGCC---GGCCGTGACTGACTCGACCACCCTGGTGCCGCTCCGCGAGGACGGCACCCA 3934401

Qy        533 CGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTC 592
              |  |     || ||| |||||||   |  | ||    |  |  |  |  |    || |
Db    3934400 CCCGCCGCTGTACTGCGTGCACCCGGTCTCCGGTTCGGGCTACGTGTACGCGGGGCTCGC 3934341

Qy        593 TGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCT 652
              |     |||||| |  |  | ||||  | || |||| |||| | |||  ||||||
Db    3934340 CGAACTCCTCGGCCCTCAGCAGCCGTCGTACGGCTTCGAGGCGCCGGGGTTCGACGACGG 3934281

Qy        653 CGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCA 712
              ||    |||| |||     |||||| | ||||||  |   ||  | || | |||||||
Db    3934280 CGGGGCGCCCCTCAACTCGATCGAGGCGCTCGCGGAACGGCACCTGGCGTCCCTGCGGTC 3934221

Qy        713 GCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTT 772
              |     | ||| | || || ||| |  ||   ||   |||   ||||||  |  ||||||
Db    3934220 GGCCCGGCCGCGGGGACCGTACCTGCTGCTCGGCTGGTCGCTCGGCGGTGTGGTCGCCTA 3934161

Qy        773 CGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGA 832
              | | ||||||| || |||| |    |  || |||   ||  | | ||| ||| |  ||||
Db    3934160 CCACATGGCCCAGCGGCTCGTGGAGGCCGGCGAGGAGGTGCCGCTGCTCGTCCTCATCGA 3934101

Qy        833 CACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCT 892
```

```
Db    3934100 CGCGACGGTGCCCCAGCCGTTCCCGCTGCCGCCGGAGAAGGCGCTGCTGCACGCCTTCCT 3934041

Qy        893 GCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGACGGAGACACGGCGAT 952

Db    3934040 GCGGGACCTCGTGACGGCGATGGGGGAGCCGCCGGAGGCGCTGGACGCGCTGCTGACGCC 3933981

Qy        953 CAAGCCCGACGACCTCGCC 971

Db    3933980 CTACCCCGACGACGTCACC 3933962


RESULT 3
US-10-156-761-5999
; Sequence 5999, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5999
;   LENGTH: 1257
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1257)
US-10-156-761-5999

  Query Match           11.2%;  Score 112.4;  DB 6;  Length 1257;
  Best Local Similarity   49.0%;  Pred. No. 6.2e-17;
  Matches  361;  Conservative    0;  Mismatches  366;  Indels    9;  Gaps    2;

Qy        219 GCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCC 278

Db        106 GCCGCGCCCGCGCCCGCCGCCGCCCCGGCCGCCCCGGCTCCCGCCCCGGCTGCCCCGGCT 165

Qy        279 CCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCG 338

Db        166 CCGGCCGCTCCGGCTCCGGCTCCGGCCGCCCCGGCCCCGGCTCAGGCCGCCCCCGCGGCG 225

Qy        339 GCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCCG 398

Db        226 CCCGCCGCCCCGGCTCAGGCCGCTCCGGCCCCGGCCGCTCCGCAGGCCCCGGCCGCCCCG 285

Qy        399 GCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGCG 458

Db        286 GCGCCGGCCCCCGTCACCCCGGCCGCGGCGCCCGCCGCGACCTCCGGTGACGACGGCGCC 345

Qy        459 GACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAAC 518

Db        346 TACGTGACCCCGCTGGTGCGCAAGCTCGCCGCCGAGAACGGCGTCGACCTGGCCACCGTC 405

Qy        519 CCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCAGCCG 578

Db        406 AAGGGCACCGGCGTCGGCGGCCGCATCCGCAAGCAGGACGTCACCGCCGCCGCCGCGAGGCC 465

Qy        579 CTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGCCCGG 638

Db        466 GCGAAGGCCGCCGCTCCGGCGCCCGCCGCCGCT---CCGGCCGCCGCGAAGAAGGCCCCG 522

Qy        639 GGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTACGTC 698

Db        523 GCCCTGGAGGTCTCCCCCCTCCGCGGCCAGACCGTCAAGATGCCCCGCATCCGCAAGGTC 582

Qy        699 GCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTCCGGC 758

Db        583 ATCGGC------GACAACATGGTGAAGGCCCTGCACGAGCAGGCCCAGCTCTCCTCGGTC 636

Qy        759 GGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTCCCGG 818

Db        637 GTCGAGGTGGACATCACCCGCCTGATGAAGCTCCGCGCGCAGGCGAAGGACTCCTTCGCG 696

Qy        819 CTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGGCGAC 878

Db        697 GCCCGCGAGGGCGTCAAGCTCTCCCCGATGCCGTTCTTCGTGAAGGCGGCGGCCCAGGCG 756

Qy        879 ATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCGCGAC 938

Db        757 CTGAAGGCCCACCCGGCCGTCAACGCCCGGATCAACGTGGACGAGGGCACGATCACCTAC 816

Qy        939 GGAGACACGGCGATCA 954

Db        817 TTCGACACCGAGAACA 832


RESULT 4
US-10-282-122A-13806
; Sequence 13806, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13806
;   LENGTH: 1121
;   TYPE: DNA
```

```
;   ORGANISM: Burkholderia mallei
US-10-282-122A-13806

  Query Match             11.0%;  Score 110;  DB 7;  Length 1121;
  Best Local Similarity   48.8%;  Pred. No. 2.3e-16;
  Matches  296;  Conservative    0;  Mismatches  310;  Indels    0;  Gaps    0;

Qy        308 CGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGC 367
              ||  || | ||   |   ||| | |   || | |    |   ||      || ||  ||  |
Db         63 CGACGCGCTGGGCGTGACCGTACCGGTCGAGTCGATCCTCGCGAGTCCGACCATCGTCGA 122

Qy        368 CCCCGCGGCCTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGA 427
              |  || |  | ||  |           | ||  ||   ||| ||| |   | |||||  |
Db        123 GCTTGCCGGATACGTCGAGCAATTGCGCTCGCGGGACGTCCGCGACGGCGCTGCGCCCGT 182

Qy        428 TCCGCGCGCGGGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGG 487
              ||  ||  |     |  |  || ||  |   |       |||||  | ||  |||
Db        183 GTCGATCGGCGAAACACCGGCGGACGCGGATGCTCGAGCTCAGGCGCAGGCGGATACGGT 242

Qy        488 GCCCAACCCCGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTG 547
              || || |      |  | ||||| |           ||| |  ||  | |      ||||||
Db        243 GCACACCGATTGCCTGATCGTCATTCAAGCAGGCGGCGCCGAACAAGCGCCGGTGTTCTG 302

Qy        548 GGTGCACGGCGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCAC 607
              | | ||||||   ||    ||    | || ||  | | ||  |   |  |||  || |
Db        303 CATCCCGGGCGCGGGGGGCAGCGTCGCGTCGTTCGTTGCGCTTGCGAGCATGCTGCGCGC 362

Qy        608 GGATTACCCGTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCAC 667
              ||    ||| | | || ||| ||| | || ||  | ||||| || |    |||   |
Db        363 CGACATACCGGTATACGGCTTGCAGCCTCGCGGGCTGGACGGCCTGGGGCCGCCGGACCG 422

Qy        668 GAGCATCGAGGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGG 727
              | |||| | |   ||| ||  |||| |  | | | |   | |||   ||
Db        423 ATCCGTCGAAGCGGCTGCGCGCCGGTACGCGCGGGCCATTCTGGATGCCACCCCGCCCGG 482

Qy        728 GCCTTACGTCGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCA 787
              |||       | | |  || |  || | |||||||   |||||  ||||||  || ||||
Db        483 GCCGCCGCGCATCGTCGGCCACTCGTTCGGCGGCTGGATCGCGCTCGAGACAGCGCGGCT 542

Qy        788 GCTCCTCTCGCAAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGT 847
              |||   |      ||   ||||  | ||| ||| ||| |||||||  |  ||||  ||
Db        543 GCTGGACGGCATGGGAACGCGCTGCGCCCCGCTCGTCCTGCTCGATTCGAATCCACCGCC 602

Qy        848 CGCGGGCCTCGCGCAGGAGACGCCGGGCGACATCGACCCGATCCTGCCGCTCCTGCTCAT 907
              ||||   |  ||   |   ||||  |||  |||  | | |  ||||| |   | |
Db        603 CGCGTCGCAGGCCTGGCGCGCGCCTTCCGAGGCAGACATGCTGCGCACGCTCGTCGGCCT 662

Qy        908 GGCCAA 913
              |  | |
Db        663 GCTCGA 668


RESULT 5
US-10-211-028-3
; Sequence 3, Application US/10211028
; Publication No. US20050027113A1
; GENERAL INFORMATION:
;  APPLICANT: CUBIST PHARMACEUTICALS, INC.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO THE DAPTOMYCIN
;  TITLE OF INVENTION:  BIOSYNTHETIC GENE CLUSTER
;  FILE REFERENCE: CUB-12 PCT CIP
;  CURRENT APPLICATION NUMBER: US/10/211,028
;  CURRENT FILING DATE:  2002-07-31
;  PRIOR APPLICATION NUMBER: PCT/US02/24310
;  PRIOR FILING DATE: 2002-10-25
;  PRIOR APPLICATION NUMBER: PCT/US01/32354
;  PRIOR FILING DATE: 2001-10-17
;  PRIOR APPLICATION NUMBER: 60/310,385
;  PRIOR FILING DATE: 2001-08-06
;  PRIOR APPLICATION NUMBER: 60/379,866
;  PRIOR FILING DATE: 2002-05-10
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3
;   LENGTH: 7140
;   TYPE: DNA
;   ORGANISM: Streptomyces roseosporus
US-10-211-028-3

  Query Match             10.8%;  Score 107.8;  DB 8;  Length 7140;
  Best Local Similarity   47.6%;  Pred. No. 5.5e-16;
  Matches  445;  Conservative    0;  Mismatches  482;  Indels    8;  Gaps    4;

Qy         42 GACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCC 101
              ||||    ||| ||| ||  ||||  |   |||| ||  |     || ||    | ||| |
Db       5890 GACGGGCTCGGCGACCAGCGGCTCGTCGCGTACGTCACCGGCGGCC-CCGGCACACCGTC 5948

Qy        102 TTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACC 161
              |   ||  | | ||  |  |    ||   ||   |  |||      |  || |
Db       5949 GGCCGCCGCGCTGCGCGCCCACCTGGCGTCCCGGCTGCCCCGGCACATGGTGCCCGGTGA 6008

Qy        162 CATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCA 221
              | |||  |    |  |  | || |||   ||  | |    |   |||||||
Db       6009 CGTCCTCACCCTGGACGCCCTGCCGCTCACCGCCAACGGCAAGGTGGACCGCACGGCGCT 6068

Qy        222 GCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCG 281
              ||||||| || |  | |  | || | ||  | |  |     |  | ||||| | |   ||
Db       6069 GCCCGGCCCCGGCACCCAGACCGCCGCCCCCGGGCGCGCACCCCAGTCGCCGCAGGAACG 6128

Qy        282 GCGGC--CTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGG 339
              |  ||    |  |      || || | | ||||  |    | || ||      |  ||
Db       6129 GGTGCTGTGCGCCTTGTTCGCCGACGTGCTCGGCCGGGAGACCGTCGGCGTGGACGAGGG 6188

Qy        340 CCTCCGCAGCCCCGGCGG--CCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
              ||  |  ||  |||||   |||   |  | || |   |  |||||||   |   | |
Db       6189 GTTCTTCGACCTGGGCGGTCACTCGCTGCTCGCCACTCGCCTCGCGGCCCGGGTCCGCGC 6248

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
              ||||  |   |||     | | ||    | |  || || |   ||  |   ||  |
Db       6249 GGCGCTGGGCGTGGAGATCTCCGTGCGCACCCTGTTCGAGGCGCCGACCCCTGCCCTGCT 6308

Qy        458 GGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
              |          | ||  |  ||   ||| |  | ||||   ||||   ||  |
Db       6309 CGCGTCGGCGTGCACGGCGGACGCCGCGGCGTACGACCCGTTCGAGACGGTGCTGCCGCT 6368

Qy        518 CCCGCGCGGCTCACGCGGGAGCTC---GTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCA 574
              | ||||       || ||    |  |||||| || |||| ||      || ||
Db       6369 GCGGCGCACGGGCAGCCGGCCACCGCTGTTCTGCGTCCACGCCGGAATGGGCCTGAGCTG 6428

Qy        575 GCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGC 634
              | ||  |   ||||     |  || | | ||||    ||| | | ||   | |||||
Db       6429 GGCGTACGCCGGCCTGCTCAGCCATCTGGACGCGGACGTGCCGGTTTACGGACTGCAGGC 6488

Qy        635 CCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTA 694
              |||| |  ||  || || || |  ||     | |||| ||||||||||| || |   | ||
Db       6489 CCGGAGGCTCACCGCGCCCGGCGGGCTGCCCGGGAGCGTCGAGGAGATGGCTGAGGACTA 6548

Qy        695 CGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTC 754
              || ||      | ||||  |   | |||  ||||| |||     ||   ||  |||||
Db       6549 CGCCGGTGAGATCCGGCGCCTGTGCCCGGATGGGCCGTACCGGCTGCTCGGCTGGTCCTT 6608

Qy        755 CGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTC 814
              ||||||||   |||||  ||    | ||      |  || |    ||| ||  |  ||||
Db       6609 CGGCGGCACGGTCGCCCACGCCGTCGCGACCCGCCTGCAACAGCAGGGCCACACCGTCGA 6668

Qy        815 CCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGG 874
              |  || | | | |||||| |||| ||||    || || ||    | | |    |  ||
Db       6669 ACTCCTCGCCGTCCTCGACGCCTACCCCGTCACCGGGGCCCGGCCCGACGCCGAGGTGGA 6728

Qy        875 CGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCG 934
              |||      |   ||      |||   | |  |   |||  | ||  |  ||| ||
Db       6729 CGAACAGCGCATCGTCGCCGACTACCTCGCCCAGCTCGGTTCCCCCGTCGCCCCCGAGCG 6788

Qy        935 CGACGGAGACACGGCGATCAAGCCCGACGACCTCG 969
              |  ||  | |   |||       || ||   |||||
Db       6789 CCTCGAGGGCGACGCGTGGCTCCCGGAGTTCCTCG 6823
```

```
RESULT 6
US-10-211-028-1
; Sequence 1, Application US/10211028
; Publication No. US20050027113A1
; GENERAL INFORMATION:
;  APPLICANT: CUBIST PHARMACEUTICALS, INC.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO THE DAPTOMYCIN
;  TITLE OF INVENTION:  BIOSYNTHETIC GENE CLUSTER
;  FILE REFERENCE: CUB-12 PCT CIP
;  CURRENT APPLICATION NUMBER: US/10/211,028
;  CURRENT FILING DATE:  2002-07-31
;  PRIOR APPLICATION NUMBER: PCT/US02/24310
;  PRIOR FILING DATE: 2002-10-25
;  PRIOR APPLICATION NUMBER: PCT/US01/32354
;  PRIOR FILING DATE: 2001-10-17
;  PRIOR APPLICATION NUMBER: 60/310,385
;  PRIOR FILING DATE: 2001-08-06
;  PRIOR APPLICATION NUMBER: 60/379,866
;  PRIOR FILING DATE: 2002-05-10
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 90597
;   TYPE: DNA
;   ORGANISM: Streptomyces roseosporus
US-10-211-028-1

  Query Match             10.8%;  Score 107.8;  DB 8;  Length 90597;
  Best Local Similarity   47.6%;  Pred. No. 3.6e-16;
  Matches  445;  Conservative    0;  Mismatches  482;  Indels    8;  Gaps    4;

Qy         42 GACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTCTCCGCCGCC 101
              ||||    ||| ||| ||  ||||  |   |||| ||  |    || ||    | ||| |
Db      83946 GACGGGCTCGGCGACCAGCGGCTCGTCGCGTACGTCACCGGCGGCC-CCGGCACACCGTC 84004

Qy        102 TTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCCGCGCACACC 161
                |   ||  | | ||  |  |    ||   ||   |  |||       |  || |
Db      84005 GGCCGCCGCGCTGCGCGCCCACCTGGCGTCCCGGCTGCCCCGGCACATGGTGCCCGGTGA 84064

Qy        162 CATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGCAACGGCGCA 221
              | |||   |      |  |   | || |||   ||  | |      |     |||||||
Db      84065 CGTCCTCACCCTGGACGCCCTGCCGCTCACCGCCAACGGCAAGGTGGACCGCACGGCGCT 84124

Qy        222 GCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCCACGGCCCCG 281
              ||||||| || |  | | |  | || | ||  | |      |  | ||||| | |   ||
Db      84125 GCCCGGCCCCGGCACCCAGACCGCCGCCCCCGGGCGCGCACCCCAGTCGCCGCAGGAACG 84184

Qy        282 GCGGC--CTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCCCCCGCGG 339
              |  ||    | |     || || | | |||||  |    | ||| ||      |   ||
Db      84185 GGTGCTGTGCGCCTTGTTCGCCGACGTGCTCGGCCGGGAGACCGTCGGCGTGGACGAGGG 84244

Qy        340 CCTCCGCAGCCCCGGCGG--CCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCAGTCCC 397
                ||  |  ||  |||||   |||   |  | || |   |  |||||||   |    | |
Db      84245 GTTCTTCGACCTGGGCGGTCACTCGCTGCTCGCCACTCGCCTCGCGGCCCGGGTCCGCGC 84304

Qy        398 GGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCCGGGGC 457
              ||||  |  |   |||     | | ||     | |  || || |    |  |   || |
Db      84305 GGCGCTGGGCGTGGAGATCTCCGTGCGCACCCTGTTCGAGGCGCCGACCCCTGCCCTGCT 84364

Qy        458 GGACAGCATCGCCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCTCGTCGTCATCAA 517
              |     |  | ||  |  ||   ||| |  | ||||   ||||   ||  |
Db      84365 CGCGTCGGCGTGCACGGCGGACGCCGCGGCGTACGACCCGTTCGAGACGGTGCTGCCGCT 84424

Qy        518 CCCGCGCGGCTCACGCGGGAGCTC---GTTCTGGGTGCACGGCGCGCCTGGGCTCGCGCA 574
              | ||||       || ||    |  |||||| || |||| ||     || ||
Db      84425 GCGGCGCACGGGCAGCCGGCCACCGCTGTTCTGCGTCCACGCCGGAATGGGCCTGAGCTG 84484

Qy        575 GCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGGC 634
              | ||  |    ||||    |  || | | ||||   ||| | | ||   | |||||
Db      84485 GGCGTACGCCGGCCTGCTCAGCCATCTGGACGCGGACGTGCCGGTTTACGGACTGCAGGC 84544

Qy        635 CCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATTA 694
              |||| |  ||  || || || | ||     |  |||| |||||||||| || |   | ||
Db      84545 CCGGAGGCTCACCGCGCCCGGCGGGCTGCCCGGGAGCGTCGAGGAGATGGCTGAGGACTA 84604

Qy        695 CGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCTC 754
              || ||      | ||||  |   | |||   ||||| |||    ||   ||   |||||
Db      84605 CGCCGGTGAGATCCGGCGCCTGTGCCCGGATGGGCCGTACCGGCTGCTCGGCTGGTCCTT 84664

Qy        755 CGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCTC 814
              ||||||||   |||||  ||   | ||   |  || |    ||| ||  |   ||||
Db      84665 CGGCGGCACGGTCGCCCACGCCGTCGCGACCCGCCTGCAACAGCAGGGCCACACCGTCGA 84724

Qy        815 CCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGGG 874
              |  || | | | |||||| |||| ||||    || || ||   |  | |    |  ||
Db      84725 ACTCCTCGCCGTCCTCGACGCCTACCCCGTCACCGGGGCCCGGCCCGACGCCGAGGTGGA 84784

Qy        875 CGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATCG 934
              |||       |   ||         |||   | | |  |||  | ||  |  ||| ||
Db      84785 CGAACAGCGCATCGTCGCCGACTACCTCGCCCAGCTCGGTTCCCCCGTCGCCCCCGAGCG 84844

Qy        935 CGACGGAGACACGGCGATCAAGCCCGACGACCTCG 969
              |  ||  | |   |||      || ||  |||||
Db      84845 CCTCGAGGGCGACGCGTGGCTCCCGGAGTTCCTCG 84879


RESULT 7
US-10-389-647-202
; Sequence 202, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
;  APPLICANT: GREENBERG, E. Peter
;  APPLICANT:  SCHUSTER, Martin
;  APPLICANT:  LOSTROH, Candi
;  TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
;  FILE REFERENCE: UIZ-038CP
;  CURRENT APPLICATION NUMBER: US/10/389,647
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 09/653730
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/153022
;  PRIOR FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 710
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 202
;   LENGTH: 7059
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-10-389-647-202

  Query Match             10.6%;  Score 106.4;  DB 7;  Length 7059;
  Best Local Similarity   50.8%;  Pred. No. 1.2e-15;
  Matches  308;  Conservative    0;  Mismatches  291;  Indels    7;  Gaps    2;

Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||           ||   ||
Db       5956 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6015

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| ||||| |  |  |  |  || | ||  |  |   ||| |||      ||| |
Db       6016 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6075

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
              | |    ||  |  | |  ||||    |||   |||  | | |  | |  ||
Db       6076 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6131

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
              | |    |  |  ||  || |     |  | | | | | | ||| || |  |
Db       6132 GACGCCGACCGTGGAGGGCCTGGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6191

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              |||| ||| ||    |||  | ||||  ||| |     |||| |    || |||
Db       6192 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6251
```

```
Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
              || | || ||       |   || |||     ||| ||     ||||  | ||
Db       6252 GCTCGGCGGCCACGTGCTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6311

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | ||||    |||  || |  |  |   | |||  |  ||  | ||||
Db       6312 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6371

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              |||  ||||| |||  |||| ||||   | | || | |       |||  ||| || ||
Db       6372 GGACATCGCCGCCAGTTACCTCGCGGCCATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6431

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| ||||||||||||||| || | | |
Db       6432 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6491

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |   | ||  ||| |||  ||| || ||||||||  ||    |||  | |
Db       6492 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6551

Qy        854 CCTCGC 859
              || |||
Db       6552 CCACGC 6557
```

RESULT 8
US-11-014-621-26
; Sequence 26, Application US/11014621
; Publication No. US20050227345A1
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Lee, Kimberly
; APPLICANT: Greenberg, E. P.
; APPLICANT: Muh, Ute
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038
; CURRENT APPLICATION NUMBER: US/11/014,621
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: US/09/653,730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153,022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 7110
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-11-014-621-26

Query Match 10.6%; Score 106.4; DB 10; Length 7110;
Best Local Similarity 50.8%; Pred. No. 1.2e-15;
Matches 308; Conservative 0; Mismatches 291; Indels 7; Gaps 2;

```
Qy        257 CTCGCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCC 316
              | || || | |    | |  ||  | ||| |||| | ||          ||   ||
Db       6007 CGCGACGACTACGAGCGCACCCTGGCCGGACTCCTCGGCGAGTTGCTGGATCGTCCCCGG 6066

Qy        317 GGCGGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGC 376
              |  ||| |||||  |  |  |  |  || | ||  |  |   ||| |||     |||  |
Db       6067 GTAGGCATCCGCGACAGCTTCTTCGACCTCGGCGGCACCTCGCTCAGCGCGATGCGCTTC 6126

Qy        377 CTCCGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGC 436
               | |       ||   |  | |   ||||      |||    |||  | | |  | |  ||
Db       6127 ATGCTGCTGATCGAGAAGCGCTATGGCG----TCGACCTGCCGATGGCCGCGCTGATCGA 6182

Qy        437 GGGGGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCC 496
              |  |    |  |  ||  || |    |  | | | |  | ||| || |  |
Db       6183 GACGCCGACCGTGGAGGGCCTGGCCGAACGCCTGCGGGAACGCTCGGCGGTGCGCGCCTT 6242

Qy        497 CGACGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGG 556
              ||||  ||| ||    |||    |  ||||   |||  |     ||||   || |||
Db       6243 CGACCCGCTGGTACCGATCCGTGCCGGCGGCAGCCGCCCGCCGCTGTTCCTCGTCCACCC 6302
```

```
Qy        557 CGCGCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCC 616
              || | || ||       |   || |||     ||| ||     ||||  | ||
Db       6303 GCTCGGCGGCCACGTGCTCTGCTACCTGCCGCTGGTCCGCGCACTGCCGCCGGACCAGCC 6362

Qy        617 GTTCTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGA 676
              | | |  ||| | |||||    |||  || |  |  |   | |||  |  ||  | ||||
Db       6363 GGTATATGCCCTGCAGGCGGCCGGCACCGGCCAGGGCAGTACGCCGCTGGCGGTCCTCGA 6422

Qy        677 GGAGATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGT 736
              |||  ||||| |||  |||| ||||   | | || | |       |||  ||| || ||
Db       6423 GGACATCGCCGCCAGTTACCTCGCGGCCATCCGCCGGGTGCAGCCGGAAGGCCCCTATTA 6482

Qy        737 CGTGGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTC 796
              | | || ||   ||| | ||||||| || |||||| ||||||||||||||| || | | |
Db       6483 CCTCGGCGGCTGGTCGTTCGGCGGCTTCGTCGCCTACGAGATGGCCCGGCAACTGCGCGC 6542

Qy        797 GC---AAGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGG 853
              ||   |   | ||  ||| |||  ||| || ||||||||  ||    |||  | |
Db       6543 GCTCGACCCGCAGGCGGTCGCCCAACTGATCGTGCTCGACTCCATCACCGTCGACCGCAA 6602

Qy        854 CCTCGC 859
              || |||
Db       6603 CCACGC 6608
```

RESULT 9
US-10-132-134-15
; Sequence 15, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 5862
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-15

Query Match 10.3%; Score 102.8; DB 6; Length 5862;
Best Local Similarity 52.4%; Pred. No. 8.3e-15;
Matches 302; Conservative 0; Mismatches 262; Indels 12; Gaps 3;

```
Qy        260 GCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGC 319
              || ||   |||||  ||| | | ||  |  ||  ||  |     ||  ||  | |   ||
Db       4668 GCAGTTCCTCGCCCAGGCGCAGACGCACCTCGACGCGTCCCTCACCCTCGACGTCATCGC 4727

Qy        320 GGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTC 379
              | || || | |    || |  ||  | | ||    ||  | | ||| | |   ||
Db       4728 CGACTGCGAAACGCTGCAGGACGTCGTCGACGCGATCACCGGCACCGCCTCTGACACCGG 4787

Qy        380 CGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGG 439
              | ||||| || ||   || ||  ||       | |||||| |     ||  | |  ||||
Db       4788 CACGGCCGCCCCAAAGCCCGCTTCG------GTGGCTCCCGTTGAGGAGGCCCCGGCGGC 4841

Qy        440 GGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGA 499
              | ||||| | |   |||                | |    | |||   ||  |  | || |
Db       4842 CGCCGCACCGGCAAAGGCCCCGGTACGGCGCGCCGCGGCCGCATCGCCGAACGACTTCCC 4901

Qy        500 CGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGC 559
              ||| || |||  ||| |||      |  |||  || ||  |  |||||||| |||  ||
Db       4902 CGAACTGGTCCGCATGAACGGTGTCACCTCCGGCCGGCCCGTGTTCTGGGTCCACCACGG 4961

Qy        560 GCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTT 619
                ||   || | || |   |||  ||| |   ||||| | | || ||  |   || ||
```

```
Db       4962 CAACGGCGGCGTGGAGTC---CTACGCCCCGCTGGCCGCACGCTGCCCGCGT---CCCTT 5015

Qy        620 CTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGA 679

Db       5016 CTACGGCATCCAGCCGAAGGGCTGGATCGACTCCACCGACATCCTCACCGGTCAGTACGC 5075

Qy        680 GATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGT 739

Db       5076 CATGGCCGAGCACTACGCGTCCCTCATCCTCGCCGTACAGCCGGAAGGCCCGTACGACAT 5135

Qy        740 GGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCA 799

Db       5136 CGGCGGGTTCTCCCTGGGCGGCCTGTTCGCGTACGAGACCGTGCGGCAGCTCCAGCTGAC 5195

Qy        800 AGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACAC 835

Db       5196 GGGCGCCGACGTGCGCACGCTGGTCATGCTGGACAC 5231


RESULT 10
US-10-132-134-1
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Yang, Xianshu
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
;  FILE REFERENCE: 3012-2US
;  CURRENT APPLICATION NUMBER: US/10/132,134
;  CURRENT FILING DATE:  2002-04-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 52101
;   TYPE: DNA
;   ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-1

  Query Match            10.3%;  Score 102.8;  DB 6;  Length 52101;
  Best Local Similarity   52.4%;  Pred. No. 5.8e-15;
  Matches  302;  Conservative    0;  Mismatches  262;  Indels   12;  Gaps    3;

Qy        260 GCCGTCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGGC 319

Db      46439 GCAGTTCCTCGCCCAGGCGCAGACGCACCTCGACGCGTCCCTCACCCTCGACGTCATCGC 46498

Qy        320 GGCCTCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTC 379

Db      46499 CGACTGCGAAACGCTGCAGGACGTCGTCGACGCGATCACCGGCACCGCCTCTGACACCGG 46558

Qy        380 CGCGGCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGG 439

Db      46559 CACGGCCGCCCCAAAGCCCGCTTCG------GTGGCTCCCGTTGAGGAGGCCCCGGCGGC 46612

Qy        440 GGACGCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGA 499

Db      46613 CGCCGCACCGGCAAAGGCCCCGGTACGGCGCGCCGCGGCCGCATCGCCGAACGACTTCCC 46672

Qy        500 CGAGCTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGC 559

Db      46673 CGAACTGGTCCGCATGAACGGTGTCACCTCCGGCCGGCCCGTGTTCTGGGTCCACCACGG 46732

Qy        560 GCCTGGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTT 619

Db      46733 CAACGGCGGCGTGGAGTC---CTACGCCCCGCTGGCCGCACGCTGCCCGCGT---CCCTT 46786

Qy        620 CTTCGCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGA 679

Db      46787 CTACGGCATCCAGCCGAAGGGCTGGATCGACTCCACCGACATCCTCACCGGTCAGTACGC 46846

Qy        680 GATCGCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGT 739

Db      46847 CATGGCCGAGCACTACGCGTCCCTCATCCTCGCCGTACAGCCGGAAGGCCCGTACGACAT 46906

Qy        740 GGGTGGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCA 799

Db      46907 CGGCGGGTTCTCCCTGGGCGGCCTGTTCGCGTACGAGACCGTGCGGCAGCTCCAGCTGAC 46966

Qy        800 AGGGGAGCGCGTCTCCCGGCTGGTCATGCTCGACAC 835

Db      46967 GGGCGCCGACGTGCGCACGCTGGTCATGCTGGACAC 47002


RESULT 11
US-10-425-115-17653/c
; Sequence 17653, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 17653
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(972)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_116102C.1
US-10-425-115-17653

  Query Match            10.2%;  Score 102.2;  DB 8;  Length 972;
  Best Local Similarity   50.1%;  Pred. No. 1.5e-14;
  Matches  276;  Conservative    0;  Mismatches  272;  Indels    3;  Gaps    1;

Qy         86 CCGCCTCTCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCT 145

Db        938 CCGCACACCCGCGCCCANAACCGCANCGGCCCCCAAGNAACGCGGGACCCCCCCGCGCCC 879

Qy        146 CAGCGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCT 205

Db        878 CCCGCCGGCGCCGCCACCCCCCGCCCGCCGCCCCCCCACCCCCCCCCCCCCCGCCACCCCC 819

Qy        206 GAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTC 265

Db        818 ACCCCCGCCCGCGCCCCCGGCCCGCCCCCCCCCCGCCCCCGCCCCCCCGCCCCCCGCCC 759

Qy        266 GATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTC 325

Db        758 ACCCCCCCCGCCCCCCGCCGCCCCCCGCCCCCCCGCCCCCCGCCCCCCGCCCCCCCCCCG 699

Qy        326 CGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGC 385

Db        698 CCGCCCCGCCCCCCCCCGCCCCCCCCCGCCGCCCACCGCGCCCCCCCCGCCCCCCCCCCC 639

Qy        386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGC 445

Db        638 CCCCCGCGCCCCGCCCCGCCCCGACCCGGCGCCCGCGCGCGA---GCGCGCGGGCGAGGC 582

Qy        446 ACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCT 505

Db        581 CGGCGCCGCCGCCCGCGGCCCGGCGCCGCCGCGCCGCCGCGGCCCCGCGCCCGCGCCCCG 522

Qy        506 CGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGG 565
```

```
Db        521 GAGGGCGCCGGCCGCGCCCGCGGCGCCCCGGCGGCGGGGCGCGGGGGAGGGCGCGCGCGC 462

Qy        566 GCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGC 625

Db        461 GCCCGCCGCCCCCCCCCCCCCGCCCCCCCCCCCNCCCCCCCCGCCCCCCCCCCCCCCCC 402

Qy        626 CTTCCAGGCCC 636

Db        401 CCCCCGCCCCC 391


RESULT 12
US-10-156-761-3145
; Sequence 3145, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3145
;   LENGTH: 810
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(810)
US-10-156-761-3145

  Query Match             10.1%;  Score 101.2;  DB 6;  Length 810;
  Best Local Similarity   51.3%;  Pred. No. 2.7e-14;
  Matches  235;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        514 TCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGGGCTCGCGC 573

Db          8 TCCGCGAGGACGGCACCCACCCGCCGCTGTACTGCGTGCACCCGGTCTCCGGTTCGGGCT 67

Qy        574 AGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGCCTTCCAGG 633

Db         68 ACGTGTACGCGGGGCTCGCCGAACTCCTCGGCCCTCAGCAGCCGTCGTACGGCTTCGAGG 127

Qy        634 CCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGCGGCCCATT 693

Db        128 CGCCGGGGTTCGACGACGGCGGGGCGCCCCTCAACTCGATCGAGGCGCTCGCGGAACGGC 187

Qy        694 ACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGGGCTGTCCT 753

Db        188 ACCTGGCGTCCCTGCGGTCGGCCCGGCCGCGGGGACCGTACCTGCTGCTCGGCTGGTCGC 247

Qy        754 CCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGAGCGCGTCT 813

Db        248 TCGGCGGTGTGGTCGCCTACCACATGGCCCAGCGGCTCGTGGAGGCCGGCGAGGAGGTGC 307

Qy        814 CCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGAGACGCCGG 873

Db        308 CGCTGCTCGTCCTCATCGACGCGACGGTGCCCCAGCCGTTCCCGCTGCCGCCGGAGAAGG 367

Qy        874 GCGACATCGACCCGATCCTGCCGCTCCTGCTCATGGCCAACTCCTTCATCAGCTTCGATC 933

Db        368 CGCTGCTGCACGCCTTCCTGCGGGACCTCGTGACGGCGATGGGGGAGCCGCCGGAGGCGC 427

Qy        934 GCGACGGAGACACGGCGATCAAGCCCGACGACCTCGCC 971

Db        428 TGGACGCGCTGCTGACGCCCTACCCCGACGACGTCACC 465


RESULT 13
US-10-282-122A-14471
; Sequence 14471, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14471
;   LENGTH: 10002
;   TYPE: DNA
;   ORGANISM: Burkholderia mallei
US-10-282-122A-14471

  Query Match             10.1%;  Score 100.8;  DB 7;  Length 10002;
  Best Local Similarity   47.7%;  Pred. No. 2.2e-14;
  Matches  356;  Conservative    0;  Mismatches  387;  Indels    3;  Gaps    2;

Qy        149 CGCCGCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAG 208

Db       7986 CGCCGCGTACGTGCGGCTGCCGGCGATGCCGCTCACGCCCAACGGCAAGCTCGAGCGCCG 8045

Qy        209 CAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGAT 268

Db       8046 CGCGCTGCCCGCCCCCGACGAGCGATCGTACGTGCGGCGCGACTACGCGGCGCCGCAGGG 8105

Qy        269 CGCCACGGCCCCGGCGGCCTC-CGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCG 327

Db       8106 CGAGATCGAGACGACACTCGCGCGGATCTGGGCGGAGCTGTTCGGCATCGAGCGCGTCGG 8165

Qy        328 CAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCC--TCCGCGGCCCCCGCGGCCTCCGCGGC 385
```

```
Db       8166 CCGGCACGACGGCTTCTTCGAACTCGGCGGGCATTCGCTGCTCGCGGTGCGGATGGTCGC 8225

Qy        386 CTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGC 445

Db       8226 GCGCGTGCACGATGTGCTGGGCGTCGAGGTGCCGCTGCGCGCGCTGTTCGCCGATCCGGT 8285

Qy        446 ACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAGCT 505

Db       8286 GCTGCACGTGTTCGCGTCGGCGGTCGCGCGCGCGTCGACGCGCCAGGCGTCGTCGAATCT 8345

Qy        506 CGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCTGG 565

Db       8346 CGTCGCGTTCCGCAGCGCGGGCACGGCCGCGCCGCTCTTCTTCATTCATTCGGGGCTCGG 8405

Qy        566 GCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTCGC 625

Db       8406 CGAGATCGGCTTCGTCGGCGATCTGCTGCCCGGCATCGCGCCGGAGATTCCGGTGTACGG 8465

Qy        626 CTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATCGC 685

Db       8466 CTTCGCGGCGGTCGGTTTCCTCGCGGGCGAGACGCCGCACGCGACGATCGAGGAGATGGC 8525

Qy        686 GGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGTGG 745

Db       8526 CGCGCAATATGTCGACGCGATGCGGCGCGTGCAGCCGCATGGGCCGTATCGGCTCGCCGG 8585

Qy        746 GCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGGGA 805

Db       8586 GTGGTGCGCGGGCGGCAACATCGCGTTCGAAATGGCCCATCAGCTGATCGCGGCCGACGA 8645

Qy        806 GCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCGGGCCTCGCGCAGGA 865

Db       8646 GACGGTCGAGTTCCTCTGCATGATCGATTCGCCGACATCCGCGCCGATCGACCGCTCGGT 8705

Qy        866 GACGCCGGGCGACATCGACCCGATCC 891

Db       8706 CACCGCGTGCGTGCTCGCGCGCATTC 8731
```

RESULT 14
US-10-132-134-37
; Sequence 37, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 37
; LENGTH: 6297
; TYPE: DNA
; ORGANISM: Streptomyces amphibiosporus
US-10-132-134-37

Query Match 9.8%; Score 98; DB 6; Length 6297;
Best Local Similarity 48.1%; Pred. No. 1.1e-13;
Matches 312; Conservative 0; Mismatches 330; Indels 6; Gaps 1;

```
Qy        204 CTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCG 263

Db       5056 CTACGCGAGTACGGACTCGACTCGATCTTCCTCACCCAGTTCCTCGGCACGGCCCGCAAG 5115

Qy        264 TCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCC 323

Db       5116 GAGTTCGACCCGGCGCTCACGCTCGACGTCATAGCGGGCTGCGAGACGCTGACGGACTTC 5175

Qy        324 TCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCG 383

Db       5176 ATCGACGCGATCGAGCGCGCCGTCGCCCCGCCGGCCGCCACGCCTCCGGCGCCCGCATCC 5235

Qy        384 GCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGAC 443

Db       5236 GCGTCCGCGCCCTCACCCTCGTCCGGTACGGAAGGGGAACCGTCAACGCGCCCCGCCCCG 5295

Qy        444 GCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAG 503

Db       5296 GAGCCGGATGGCGTACCGGTGGTGCGTCCCGTCGCCAAGGCTCCCGAGGAGTTCCCCGAG 5355

Qy        504 CTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCT 563

Db       5356 CTGATCCCCATGAACGCCGTACGGGAGGGCCGCCCGGTCTTCTGGGTCCACCACGGCAAC 5415

Qy        564 GGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTC 623

Db       5416 GGCGGAGTCGAGTCGTACGCGGCGGTGGCCGAGTGCTGCGGAC------GCCCCTTCTAC 5469

Qy        624 GCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATC 683

Db       5470 GGCATCCAGCCCCGCGGCTGGACGGGCTCGGAGGACATCCTCACCGGCCAGGAGGCCATG 5529

Qy        684 GCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGT 743

Db       5530 GCCGCCTACTACGTGGACATCATCCGCGCCGTCCAGCCGGAGGGCCCGTACGACGTCGGC 5589

Qy        744 GGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGG 803

Db       5590 GGCTTCTCCCTCGGCGGTCTGTTCGCCTACGAGGTCGTACGCCAACTCCAGCTCCAGGAC 5649

Qy        804 GAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCG 851

Db       5650 GCCACGGTGGACACCCTGGTGATGCTGGACACCCTCGACGCCGCCTCG 5697
```

RESULT 15
US-10-132-134-25
; Sequence 25, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 50543
; TYPE: DNA
; ORGANISM: Streptomyces amphibiosporus
US-10-132-134-25

Query Match 9.8%; Score 98; DB 6; Length 50543;
Best Local Similarity 48.1%; Pred. No. 7.6e-14;
Matches 312; Conservative 0; Mismatches 330; Indels 6; Gaps 1;

```
Qy        204 CTGAGCAGCAACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCG 263

Db      44683 CTACGCGAGTACGGACTCGACTCGATCTTCCTCACCCAGTTCCTCGGCACGGCCCGCAAG 44742

Qy        264 TCGATCGCCACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCC 323

Db      44743 GAGTTCGACCCGGCGCTCACGCTCGACGTCATAGCGGGCTGCGAGACGCTGACGGACTTC 44802

Qy        324 TCCGCAGCCCCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCG 383

Db      44803 ATCGACGCGATCGAGCGCGCCGTCGCCCCGCCGGCCGCCACGCCTCCGGCGCCCGCATCC 44862
```

```
Qy        384 GCCTCCGCAGTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGAC 443
              || |||||   | |  |  ||  |    ||    ||  ||     ||||| | |
Db      44863 GCGTCCGCGCCCTCACCCTCGTCCGGTACGGAAGGGGAACCGTCAACGCGCCCCGCCCCG 44922

Qy        444 GCACGGCCCGGGGCGGACAGCATCGCCCCGCAGCCCGAGCCGGGGCCCAACCCCGACGAG 503
              |  | |   || |        |    ||     || || |        |   |  ||||
Db      44923 GAGCCGGATGGCGTACCGGTGGTGCGTCCCGTCGCCAAGGCTCCCGAGGAGTTCCCCGAG 44982

Qy        504 CTCGTCGTCATCAACCCGCGCGGCTCACGCGGGAGCTCGTTCTGGGTGCACGGCGCGCCT 563
              ||  ||  ||| ||| |      |      || |       |||||||| |||  ||
Db      44983 CTGATCCCCATGAACGCCGTACGGGAGGGCCGCCCGGTCTTCTGGGTCCACCACGGCAAC 45042

Qy        564 GGGCTCGCGCAGCCGCTCTATCCCCTGTCTGCCGCGCTCGGCACGGATTACCCGTTCTTC 623
              ||    |   || ||  |    |  || | |       |||          ||| |||| |
Db      45043 GGCGGAGTCGAGTCGTACGCGGCGGTGGCCGAGTGCTGCGGAC------GCCCCTTCTAC 45096

Qy        624 GCCTTCCAGGCCCGGGGCGTCGACGGGCTCGCCATGCCCTTCACGAGCATCGAGGAGATC 683
              | | ||||| |||| |||      ||     |       | ||||  ||   ||||  ||
Db      45097 GGCATCCAGCCCCGCGGCTGGACGGGCTCGGAGGACATCCTCACCGGCCAGGAGGCCATG 45156

Qy        684 GCGGCCCATTACGTCGCCTGCCTGCGGCAGCGTAGTCCGAGAGGGCCTTACGTCGTGGGT 743
              || ||| | ||||| | |  | | ||          |||   || || |||| ||| ||
Db      45157 GCCGCCTACTACGTGGACATCATCCGCGCCGTCCAGCCGGAGGGCCCGTACGACGTCGGC 45216

Qy        744 GGGCTGTCCTCCGGCGGCATCATCGCCTTCGAGATGGCCCGGCAGCTCCTCTCGCAAGGG 803
              ||  | |||  ||||||  |  |||||| |||| | |  || || ||||     || |
Db      45217 GGCTTCTCCCTCGGCGGTCTGTTCGCCTACGAGGTCGTACGCCAACTCCAGCTCCAGGAC 45276

Qy        804 GAGCGCGTCTCCCGGCTGGTCATGCTCGACACCTATCCCGCGGTCGCG 851
              |      ||    |  |||||  ||||| ||||||     ||| | | ||
Db      45277 GCCACGGTGGACACCCTGGTGATGCTGGACACCCTCGACGCCGCCTCG 45324
```

Search completed: April 27, 2006, 14:39:50
Job time : 1008.57 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 627.947 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_55000_55999
Perfect score: 1000
Sequence: 1 tcacggcgatggtctgcgat.....cccccgatgcagctcccgcg 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	91.4	9.1	88421	14	US-11-205-109-1	Sequence 1, Appli
	2	79.6	8.0	4770	7	US-10-496-351-3	Sequence 3, Appli
c	3	79.6	8.0	82746	7	US-10-496-351-56	Sequence 56, Appl
c	4	79.2	7.9	2352	14	US-11-052-554A-541	Sequence 541, App
	5	78.8	7.9	1738	9	US-10-501-035-144	Sequence 144, App
c	6	78.4	7.8	1280	8	US-10-802-796-4	Sequence 4, Appli
c	7	78.4	7.8	1280	11	US-11-221-284-4	Sequence 4, Appli
	8	78	7.8	113193	7	US-10-478-943D-1	Sequence 1, Appli
	9	77.8	7.8	2247	7	US-10-496-351-8	Sequence 8, Appli
c	10	77.2	7.7	3240	14	US-11-052-554A-529	Sequence 529, App
c	11	77	7.7	687411	9	US-10-330-773-26	Sequence 26, Appl
	12	76.6	7.7	14055	7	US-10-496-351-1	Sequence 1, Appli
c	13	76.4	7.6	1086	14	US-11-052-554A-553	Sequence 553, App
	14	76.4	7.6	1511	8	US-10-775-169-313	Sequence 313, App
	15	76	7.6	4617	14	US-11-052-554A-530	Sequence 530, App
c	16	75.6	7.6	88421	14	US-11-205-109-1	Sequence 1, Appli
c	17	75.4	7.5	12277	8	US-10-477-507A-3	Sequence 3, Appli
c	18	75.2	7.5	3990	14	US-11-052-554A-520	Sequence 520, App
	19	74.8	7.5	2989	8	US-10-947-249-114	Sequence 114, App
	20	74.6	7.5	9222	7	US-10-496-351-2	Sequence 2, Appli
	21	74.4	7.4	5706	14	US-11-052-554A-519	Sequence 519, App
	22	74.4	7.4	9975	7	US-10-496-351-4	Sequence 4, Appli
	23	74.2	7.4	3240	14	US-11-052-554A-529	Sequence 529, App
	24	73.8	7.4	355211	9	US-10-330-773-242	Sequence 242, App
	25	73.6	7.4	8158	9	US-10-510-941-7	Sequence 7, Appli
	26	73.4	7.3	1676	11	US-11-096-568A-26551	Sequence 26551, A
c	27	73.4	7.3	4983	14	US-11-052-554A-521	Sequence 521, App
	28	73.2	7.3	3921	14	US-11-052-554A-523	Sequence 523, App
	29	73	7.3	999	11	US-11-096-568A-9639	Sequence 9639, Ap
	30	72.6	7.3	2133	14	US-11-143-980-13	Sequence 13, Appl
c	31	72.6	7.3	116856	14	US-11-143-980-1	Sequence 1, Appli
	32	72.4	7.2	789	11	US-11-096-568A-223	Sequence 223, App
c	33	72.2	7.2	880	11	US-11-096-568A-25309	Sequence 25309, A
	34	71.6	7.2	2772	14	US-11-052-554A-531	Sequence 531, App
	35	71.4	7.1	3293	8	US-10-770-726-43	Sequence 43, Appl
c	36	71.4	7.1	48203	8	US-10-995-561-13378	Sequence 13378, A
	37	71.4	7.1	151870	8	US-10-995-561-13199	Sequence 13199, A
	38	71.2	7.1	1077	11	US-11-096-568A-9019	Sequence 9019, Ap
c	39	71	7.1	2562	14	US-11-052-554A-533	Sequence 533, App
	40	71	7.1	3354	7	US-10-496-351-12	Sequence 12, Appl
	41	71	7.1	4146	14	US-11-052-554A-522	Sequence 522, App
c	42	70.8	7.1	2008	11	US-11-096-568A-20957	Sequence 20957, A
	43	70.6	7.1	592	8	US-10-632-150-31	Sequence 31, Appl
	44	70.6	7.1	592	14	US-11-073-457-31	Sequence 31, Appl
	45	70.6	7.1	592	14	US-11-073-460-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-11-205-109-1
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; CURRENT FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 88421
; TYPE: DNA
; ORGANISM: Actinoplanes sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2077)..(3078)
; OTHER INFORMATION: ORF 1; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3118)..(4032)
; OTHER INFORMATION: ORF 2; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4038)..(5048)
; OTHER INFORMATION: ORF 3; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6665)..(5814)
; OTHER INFORMATION: ORF 4; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7703)..(6693)

; OTHER INFORMATION: ORF 5; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9464)..(8130)
; OTHER INFORMATION: ORF 6; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9691)..(10761)
; OTHER INFORMATION: ORF 7; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12751)..(10829)
; OTHER INFORMATION: ORF 8; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (13617)..(12802)
; OTHER INFORMATION: ORF 9; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (15203)..(13614)
; OTHER INFORMATION: ORF 10; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (15591)..(15863)
; OTHER INFORMATION: ORF 11; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (15880)..(19035)
; OTHER INFORMATION: ORF 12; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (19032)..(39713)
; OTHER INFORMATION: ORF 13; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39713)..(65800)
; OTHER INFORMATION: ORF 14; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65826)..(66530)
; OTHER INFORMATION: ORF 15; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (66546)..(67370)
; OTHER INFORMATION: ORF 16; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (67384)..(70059)
; OTHER INFORMATION: ORF 17; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (70099)..(70662)
; OTHER INFORMATION: ORF 18; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (70659)..(71906)
; OTHER INFORMATION: ORF 19; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (73439)..(71964)
; OTHER INFORMATION: ORF 20; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (74216)..(73563)
; OTHER INFORMATION: ORF 21; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (75424)..(74213)
; OTHER INFORMATION: ORF 22; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (78110)..(76449)
; OTHER INFORMATION: ORF 24; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (82346)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (82587)..(84446)
; OTHER INFORMATION: ORF 29; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; negative strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

Query Match 9.1%; Score 91.4; DB 14; Length 88421;
Best Local Similarity 44.8%; Pred. No. 1.9e-09;
Matches 399; Conservative 0; Mismatches 486; Indels 6; Gaps 1;

```
Qy        33 CAGGCCAGGGACGTCGACGGGGACGAGGCGCTCCGCAACTACGGCATGGATTCCCGCCTC 92
             |||| |  || ||||| | | || || |||| |  |  |  |  | |||  ||   ||
Db     64607 CAGGTCGCGGTCGTCGTCCGAGAGGACGCGCCAGGGGATAAGCGGCTGGTCGCCTATGTC 64666

Qy        93 TCCGCCGCCTTCATGCGGTCGGTGCAGCAGCGGTACGGGTCGAGCGTGCCGCTCAGCGCC 152
               || ||      |   | ||     |||| |  |    |  ||     | |     |
Db     64667 GTCGGCGGGGATGTCGAGGCGTATGCGCAGGAGCGCCTTCCCGGCTACATGGTTCCGTCG 64726

Qy       153 GCGCACACCCATCCCACCTTGAACCAGCTCACGGCCCACATTCATGGCCTCCTGAGCAGC 212
             ||   | |||||       |  | ||| || ||  ||   | | |  || | |  |
Db     64727 GCCTTCGTCCATCTGGAAGCGCTGCCGCTGACCGCGAACCAGAAGGTCGACCGGGCCGCC 64786

Qy       213 AACGGCGCAGCCCGGCACCCGTCCGCCGCGCCCCTCGCCGCGACCTCGCCGTCGATCGCC 272
               |||  ||  ||    | |  |  | ||       ||  || | ||         |
Db     64787 CTGCCCGCGCCCGAGCGGGAGACGACGACACCGGGTAAGGCACCCGCCCCCGGACCGCTC 64846

Qy       273 ACGGCCCCGGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCGGCGGCCTCCGCAGCC 332
               ||  | ||  ||      || ||||  | ||| ||  |  |    |  |   | |
Db     64847 GGCAACCTCGAGGAGTCGATGTGCCAGGCGTTCGCCGAGGTGCTCGGCCTCGACAGCGTC 64906

Qy       333 CCCGCGGCCTCCGCAGCCCCGGCGGCCTCCGCGGCCCCCGCGGCCTCCGCGGCCTCCGCA 392
               | ||| |  ||     |  ||     ||  | |  |  | |||  ||   ||
Db     64907 GGCCCGGACGACGACTTCTTCGCCCTGGGCGGCCACTCGCTGCTCGCCGTCGCGCTCGTG 64966

Qy       393 GTCCCGGCGGCGCTCCACGAGGCTCCGGCGCCTGATCCGCGCGCGGGGGACGCACGGCCC 452
               | |      |  | ||  |  | |  | | | | |      || ||| |  |
Db     64967 CAGCGGCTCAAGGCACGCGGTGTCGCCGTCACGGTGCAGGACATCATGGCCGCGCCCACG 65026
```

```
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK20155.1"
                     /db_xref="GI:13432074"
                     /translation="MITGSAHQQRGASLYARENGHFGPDLWTSAQAPVPELFLAGELL
                     HQVDGQPVGHRALQTLLHEAAHGLAHVRGIKDTSRSGNRYHNRRYAALATELGLVPPV
                     EPDGRGWSECTLTAATAVHYAEALAALADATLPYMRDPWVHVPTPVLIGAGGPAIPGT
                     DDDGDGQGGAAPTGPVVAPPPRRPGAGRAGSRLSIACGCEPPRRLSVTPRQHELAPIL
                     CGACEAPFETTDDPDLGETS"
     misc_feature    12905. .12975
                     /note="similar to GenBank Accession Number AF014839"
     CDS             complement(13201. .14616)
                     /note="similar to lambda tail fiber assembly protein and
                     Escherichia coli putative membrane protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK20156.1"
                     /db_xref="GI:13432075"
                     /translation="MTTPVQDPVAVPADPGPDPVTPPAAETLPCAWCSRPVEQRASGS
                     GRRRRYCPDRDCQDQAKRARASRRAAGGIGGDVAAAEDLLDRLETYIGRLGDGLAAEL
                     TPAGVQSAVSRAEAGAAARIAAAAAEAEASRRTAAEEVARAAAERDAALAEAQAARTT
                     AAAADETAAAARAQADADRAAAADAQRRADTAQAERERAAQDAARTAGADRDQARADAA
                     AAAADRDTAVGRAEQAETERAAARDAEQAALARAEAADRARLAAELERTAAEQARDAA
                     TARADRADEGWRAAEQARVTAQAERDAAQRAEQAAVQARDAAVADAARAWEQAADADR

  Query Match             10.8%;  Score 97;  DB 1;  Length 22437;
  Best Local Similarity   45.1%;  Pred. No. 17;
  Matches  361;  Conservative    0;  Mismatches  440;  Indels    0;  Gaps    0;

Qy         13 CGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCG 72
              ||  |  |||  |  | ||  ||   |  | |  |||    ||||    |   || |
Db      14612 CGACCCCCGTGCAGGATCCCGTCGCTGTCCCCGCCGACCCGGGCCCCGACCCCGTCACAC 14553

Qy         73 GCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCG 132
                 |   ||| | |   || |   |    | ||| ||| |    | |  |    | |
Db      14552 CGCCGGCGGCGGAGACGTTGCCGTGCGCCTGGTGCTCCCGGCCCGTGGAGCAGCGCGCCA 14493

Qy        133 TCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGG 192
              |||||||  | |    || || |  ||  ||  |      |  |||          | |
Db      14492 GCGGCTCGGGGCGGCGGCGGCGGTACTGCCCGGACCGGGACTGCCAGGACCAGGCCAAGC 14433

Qy        193 CGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCG 252
              |  |  || | ||| ||  | || |||| |  || |||        |  || || |   |
Db      14432 GGGCCCGCGCGTCGCGCCGGGCCGCCGGCGGCATCGGCGGCGACGTCGCCGCCGCCGAGG 14373

Qy        253 GCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCT 312
              || |  ||| | |  |  | ||   |  |   ||| |||||  | |  ||
Db      14372 ACCTCCTGGACCGCCTGGAGACCTACATCGGCCGGCTCGGCGACGGGCTGGCCGCCGAGC 14313

Qy        313 TCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACC 372
              |||  | | | | || | |   || | |  ||  ||||| |  |  |     || || |
Db      14312 TCACCCCCGCCGGCGTCCAGTCGGCGGTGAGCCGGGCGGAGGCCGGCGCGGCCGCCCGCA 14253

Qy        373 TCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGT 432
              |||| || | || ||||||  |  ||||    | |   | |   |  |    |  ||
Db      14252 TCGCCGCGGCCGCCGCCGAGGCTGAGGCGTCGCGGCGCACCGCGGCGGAGGAGGTGGCCC 14193

Qy        433 CACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACC 492
              |  |   ||  ||   || ||| | |  |||        |  |          |  |
Db      14192 GCGCGGCGGCCGAGCGGGACGCCGCGCTCGCCGAGGCGCAGGCGGCGCGCACCACGGCGG 14133

Qy        493 TCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGG 552
              ||||  |   |||  |  | |||||||  ||     |||  |  |         |   ||
Db      14132 CCGCCGCCGATGAGACCGCCGCGGCCGCCCGCGCCCAGGCCGACGCCGACCGGGCCGCGG 14073

Qy        553 TCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCG 612
              ||  |  ||   || || |  ||||| || || | ||   | |  |  |  | ||  ||
Db      14072 CCGCCGACGCCCAGCGCCGCGCCGACACCGCGCAGGCGGAGCGCGAGGCCGCCCAGGACG 14013

Qy        613 GCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGA 672
              || |      |  || || | |||   | | |  ||||  ||         |   | |
Db      14012 CCGCCCGCACCGCCGGCGCCGACCGGGACCAGGCCCGCGCCGACGCGGCGGCCGCCGCCG 13953

Qy        673 ACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGC 732
              |  |||| |  |   | |  | ||| ||   |    | |  | | |   ||  | | |
Db      13952 CCGACCGCGACACGGCCGTGGGCCGCGCCGAGCAGGCGGAGACCGAGCGGGCCGCGGCCC 13893

Qy        733 GCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCC 792
              |||    |||   ||||   |  ||||| || || || |||||  ||    ||   |
Db      13892 GCGACGCCGAGCAGGCCGCGCTCGCCCGCGCCGAGGCCGCCGACCGGGCACGCCTGGCCG 13833

Qy        793 TCGTCCTGCACTGGGGCGCGG 813
              ||  ||  |  |   ||| |
Db      13832 CCGAACTCGAGCGCACCGCCG 13812


RESULT 6
POL505006/c
LOCUS       POL505006              48000 bp    DNA     linear   BCT 15-APR-2005
DEFINITION  Sorangium cellulosum spiG gene (partial), spiH gene, spiI gene,
            spiJ gene, spiK gene, spiL gene and spiZ gene.
ACCESSION   AJ505006
VERSION     AJ505006.1  GI:27528333
KEYWORDS    cytochrome-P450-monooxygenase; OH-methyltransferase; polyketide
            synthase; regulatory protein; spiG gene; spiH gene; spiI gene; spiJ
            gene; spiK gene; spiL gene; spiZ gene.
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1
  AUTHORS   Knauber,J., Bloecker,H., Mueller,R., Nordsiek,G. and Beyer,S.
  TITLE     The spirangiene synthase from Sorangium cellulosum So ce90 - Module
            duplications as strategy behind the evolution of polyketide
            synthases
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 48000)
  AUTHORS   Beyer,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) Beyer S., Nbi-Mx, German Research Centre
            for Biotechnology, Mascheroder Weg 1, D-38124 Braunschweig, GERMANY
FEATURES             Location/Qualifiers
     source          1. .48000
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:56"
     gene            <1. .7558
                     /gene="spiG"
     CDS             <1. .7558
                     /gene="spiG"
                     /function="biosynthesis of spirangiene"
                     /codon_start=2
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="CAD43448.1"
                     /db_xref="GI:27528334"
                     /db_xref="GOA:Q8GBX7"
                     /db_xref="InterPro:IPR000794"
                     /db_xref="InterPro:IPR001005"
                     /db_xref="InterPro:IPR001227"
                     /db_xref="InterPro:IPR001998"
                     /db_xref="InterPro:IPR002085"
                     /db_xref="InterPro:IPR002364"
                     /db_xref="InterPro:IPR006162"
                     /db_xref="InterPro:IPR006163"
                     /db_xref="InterPro:IPR009081"
                     /db_xref="UniProt/TrEMBL:Q8GBX7"
                     /translation="ILGSDARLASALEAAGAHVTRHRDPGAWIDAIRGGGAPEVVAVL
                     ACTTRPDAPSGALAATHGAAADALAALQAWLAEERLASCALVVLTERAVATRPDEDVL
                     DLVHAPVWGLVRAAQVENPDRALILVDVDADDASLPALPALVAAAVARAADASEPQLA
                     LREGRRLVPRLGRLGGDDALVPPSAAAYRLHVAAKGTLESLALVPHEEATRPLAEGQI
                     RVAVRAAGLNFRDALDALGMYPGDPGPLGAEGAGVVLEVGPGVTELAPGDRVMGLLSA
```

AFGPIALADHRMVSRIPAGLSFVEAASIPVVFLTAYYGLVDLGRLRPGERVLVHAAAG
GVGIAAVQLARHLGAEVFATASPPKWATLRALGLDDDHIASSRSGDFEAHFRRTSGGR
GVDVVLDSLARELVDASLRLVALAPDGRFVEMGKSDVRDPLAVAASHPSVAYRAFDLA
EAGPERIQAMLAELVALFERGVLRVPPITPCDLRRAPQAFRALAQARLVGKVVLTIPR
ALDPRGSVLVTGGTGTLGALVARHLVQRHGVRSLVLSSRRGPAAPGADDLRRELEAAG
ASVAIVACDAGDRLALERLLAGIPADRPLTAVVHAAGTLDDGVLTSLTPERLRSVMSA
KVDAAAHLDELTRPLDLSAFVLFASLSGVLGSPGQANYAAANTFLDALAHQRRSRGLP
AVSLDWGYWAERTGLTAHLGDTDLQRMARGGLRPLTTDDGLALFDAALAHVDAALVAA
PFDPVALRAHAVAPILRGLVRAGAARRKLANNAASAAPLAQRLAGLAPKERERAMIDV
VRPEIGVVLGLPPSRHLDGDRPLKELGLDSLMAVELRNRLAAIVGTRLPATLLFDYPT
PTALARFLATALVGRGDAAAAPSAARAASAGASDDAIAIVAMACRFPGGVLTPEDLFR
LLREGGDAISAFPENRGWDLDALSDGAARQGGFLHDADLFDAEFFGISPRETLALDPQ
QRLLLEMTWEALERAGIIPASLHGSKSGVFVGVFGSDYVARLLHGPSGAQAMESLRGY
IGTGSASSVASGRIAYALGLEGPAISVDTACSSSLVALHLACHAIRRGECSLALAGGV
TVMATPGIFGEFSAQHAGAPDGRCKSFSADANGAGWSEGAGIVLLERLSDARKNGHPV
LAIVRGSAVNQDGKSQGLTAPNGPAQERVIRDALASAGLAPADVDAVEAHGTGTTLGD
PIEARALFATYGAAHAKERPLWLGSLKSNLGHTQAAAGIGGVIKMVLALQSGVLPRTL
HAASPSQHIDWSSGAIRLLTEEVPWKPGGSPRRAAVSSFGISGTNAHVVLEEAPRAAA
GDDQAAEPARAPLAPPVPVLLSAKSDAALRAQAGKLHAHLAEHEDVALVDVAASLATT
RSHFERRAAVVADDRAALLSALAALAEGRAGAGTVAGEALPPGKVAFVFPGQGSQWPS
MARALLASSPAFRAEIEACERALAPHVDWSLLAVLGGDEAHAAPMLERVDVVQPVLFA
VMIALAATWRAAGVTPDAVVGHSQGEIAAAYVAGALSLEDAARVVALRSRAITKLAGR
GAMAAVELTTADLEARLAPFGGRLAIAAINSPHAALVSGDPDAIDELVAELSGAQLFA
RKVRVDYASHSAHVEAIERTLLEALDGIAPRPATVPLYSAVTGERLDGEALGAAHWYR
NLRHTVRFEHATRALLDDGHRFFVEVSPHPVLTVALRDTIDASGHDAVACASLRRDEG
DLARFNLSLAELHTRGLAVDWRAFFAPFAPRTVALPTYAFQRERFWLDAPKASGVIPA
AASPAEARFWSAVEDGDVATVTDALDVGDDERRAALATLLPTLATWRRARRAQSTIDA
WRYRVAWRALTAPPKGDIAGTWLVVAPEVAGALADGVARALEARGARGVLVPVDAEDA
ARTPLAGRLRAVAGGQAIRGVISFAALDERPAPLEPSVPLGLAVSLGLVQALGDAAID
APLWLVTRGAVAVGASERVTRPLQAMTWGLGRVVALEHPERWGGLVDLGEESAPDANG
ARDANVHDRLANLVARRDDEDQLALRAAGTYVRRLVRAPHDGARPPAFAPRGAALVTG
GTGALGAHVARWLARIGAEHIVLASRRGAAAPGAAALAEELSALGARVTLAACDVADR
EAVAALVRDVEAGGATVRAVFHAGGAMHEAPVAAMRVEELADAIAAKARGAQHLHDVF
AQRPLDAFVLFSSGAGVWGGGRQGAYAAANAFLDALAEARRADGLAATSIAWGAWAGG
GMLATDAERRLKHRGVAPMDPELAVAALAHALDHAETCLAVADVDWARFAPSFASARP
RPLLDELAEARSALDALREPPDDARTAAGPEPASTLRTTLAAALPEGERHRHLLALVRT
ETAAVLGHADASRVEPNRGFFDLGLDSLMSVELRRRVQRATGIKLPATLAFDHPTPSA
LASKVLAAIVLHDATPRASPAAELERLEGMLSAIYADEALRDDLTARLRAFLDKRAVR
TERPDDAAFAEKLGSASADELIRLIDQKLGDRIDVDRY"

gene 7542. .22595
/gene="spiH"

CDS 7542. .22595
/gene="spiH"
/function="biosynthesis of spirangiene"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="polyketide synthase"
/protein_id="CAD43449.1"
/db_xref="GI:27528335"
/db_xref="GOA:Q8GBX6"
/db_xref="InterPro:IPR000794"
/db_xref="InterPro:IPR001227"
/db_xref="InterPro:IPR002198"
/db_xref="InterPro:IPR006162"
/db_xref="InterPro:IPR006163"
/db_xref="UniProt/TrEMBL:Q8GBX6"
/translation="MSTVTNDTLTEYLRRLTQELHRSETRLRATEERRHEPIAIVGLG
LRLPGGIHDRDTLWTFLEEGRDAIAPIPASRWNADATYDPDPDAVGKSYVRDAAMLDR
VDLFDADFFGISPREAKYVDPQHRLLLETSWQALEDAGIVPASLRDSKTGVFVGTGAS
DYAFLQSDRDASEAYAFMGMISSFAAGRLAFTLGLQGPALSIDTACSSSLVALHLACQ
SLRQGECDLALVAGVQVMSSPEVFVLLSRTRALASDGRSKTFSANADGYGRGEGVVVL
AVERLRDARAKGRPILAVIRGSAVNHDGTSSGITVPNGPAQQKVLRAALDDARLVPAD
VDVVECHGTGTSIGDPIEVNALAAVYGEGRPKDRPLFLGALKTNIGHLEFASGLAGVA
KMVASMRHATLPATLHTSPLNPLVDWDALPVRVVDAARPWTRRDDGAPRRAGVTAIGL
SGTNAHVIVEEAPAEPEPTTPDAAPALPAVPVLLSGKTDEALRAQAARLHAHLAGRPD
ARLVDIAASLATTRTHFDRRAAVVAADRDELLGALDALARGEAGPGSVVASAIPAGRV
VFVFPGQGSQWVGMARALLASSVVFRDEIAACERALAPHVAWSLGAVLRGDGDEATLL
GRVDVVQPVLFAVMVALAALWRSIGVTPDAVVGHSQGEIAAAYVAGALSLEDAAKVVA
LRARALTKIAGRGAMAAVELGARDTEARLAPFGDAIAIAAINSPRATLVAGDTDAIDA
LVRDLEAAQIFARKVRVDYASHSAHVEAIERELLADLAGIEPRAGAVPLYSAVTGAKL
DGSRLDAAHWFRNLRSTVRFEDATRALHDDGHRFFVEVSPHPVLALALDETLSSIGGD
AAVVSSLRRDEGDLERFLLSAGELHARGHAIDWGAFFAPLGARRVTLPTYAFQRERFW
LDAGDASDDEDAAPGASAEETAFWRAIERGDVAALSDALHVADSGRRSALESLMPALS
AWRRSRREQSTVGAWRYRVEWRPVSAASRGDVAGTWLVVAPAGVASDLASALAAALTE
RGADVVPLALEASDITAHDAASRLRDAIAGRPAIRGVLTLTPLDTAPLDGRPAVPAGL
ARSLALVQALAELAIEAPIWIVTRGAVAIGRSDRIESPVQAMAWGLGRVVSLEQPERW
GGLVDLPRAHADGALDATAIERLVAALGSREGEDELALRPTGLFARRLARAPLGDAPA
QRFEPRGTVLVTGGTGALGAHVARWLARHGAAHLVLASRRGAAAPGAEALRAELTELG
ARVTIAACDVADRAAIEELFAALPAEEPLTAIFHTAGVIDDGVLGSLTPERLETVLHG
KADAARTLHELTKTRDLAAFVLFSSAAGVFGNAGQASYAAANAFLDALAEERRALGLR
ATSVAWGAWAGAGMLTFDVAQGLRRGGFAAMEPEIAVAALASALDRDETHVTIASIDW
ARFAPSYASARQRPLVADLPDARRALDAQPSASTEHDETALVATLRALSGEARLRHLV
SLVLAQTAAVLGHADRSRLDPQRGFFDLGLDSLMAVELRRRLQKATGVKLPATITFDR
PSPHRLAGFLLDALPLGGAAPSTDARPRAARADASDEPIAIVGVGLRLPGGVDDVDAL
FRLLEGERDAVAPIPASRWDVGAVFDPDPEAKGKTYVRHAAMLDRVDLFDAGFFGISP
REARHVDPQHRLLLETAWQALEAATIVPASLRDSTTGVFVGAGASDYAVLQSSAEDAE
AYAAMGTAASFAAGRLAFTLGLQGPALSIDTACSSSLVALHLACQSLRQGECDLALAA
GVQVMASPEVFVILSRTRALAPDGRSKTFSANADGFGRGEGVVVLALERLRDARANGH
PVLAVVRGTAVNHDGASSGITAPNGSSQQKVLRAALDDARLAPTDVDVVECHGTGTSL
GDPIEVQALAAVYGEGRAPERPLRLGAVKTNIGHLEFASGLAGVVKIVASLRHGTLPA
TIHTNPRNPHIDWDALPVRVVDAAEPWERSADGSPRRAGVSAFGLSGTNAHVIVEEAP
EPDAPKAARGALPLSGALPFVLSAKSDAALRAQAAALRDQLARTPDAALVDVAAALAT
TRSQFDHRAAIAASDHGALVAALEGLAAGVAAPGTFVAKGAADKLAFLFTGQGAQRAA
MGRGLYDAFPVFRDALDAVASHLDRELDRPLRDVLFAPKGSELASLLDRTEFTQPALF
ALEVALFRLVEAWGVTPDVLLGHSVGELAAAHVAGVLSLADACTLVAARARLMQALPA
RDGAMVTVHATEAEVLAALEGRDGRAEIAAINAPSSTVIAGDVDAVLRVAAHFEARGR
KATRLRVSHAFHSHHMDPMLDAFRRVAEGLTFHPPRIPIVSNVTGRLAADEIRSPDYW
VRHVRSAVRFADGINTLEADGVSSFFELGPHGVLSALGESALANAQREVAFVPALRDG
RADVDALTAALSSLHVRGHRVDWAAFFGPFDPQTVALPTYAFQRDRFWLDAPAGATRD
EAAPAAPEDAAFWRAVDAGDVGALGATLNASGEDHLSALATLLPALSAWRRARDERSL
VDALRYRVVWKPLTPTAARDVAGTWLLVTLAGAEDALARELERALTARGAEVITMPIA
PAEADRVRLAVRVQGALADAGELRGIVSFAALDETTLATHTALPAGLALTLRLVQALG
DMGIEAPLWLVTRGAVSTGRSDRLASAAQSMTWGLGRVVGLEHPERWGGLVDVADGDT
IDARALDRLVTLLASRDGEDQLALRPTGIFARRLVRAPLGDAPPARTYTARGTSLVTG
GTGALGAHVARFLAGRGAEHLVLVSRRGEDAPGAGALRAELEALGARVTLAACDVADR
ASVAALLARLDAQGDPVRAVVHAGGVVAQAALASTDLDGVQAVVAAKVRGALHLHELL
GDRELDAFVLFASGAGVWGSGQQGAYAAGNAFLDALAEVRRAAGLTATSIAWGAWAGG
GMLADHADADAEVELRKRGLVPMAPALAIAALAGALDHGETAVTVASVDWARFAPAFA
SARPRPLLLDLPEAQRAIEAPSAPAGADGASADRALVEALRPLSEADRARHLVELVVA
ETAAVLGHADASAIDPHKGFFDLGLDSLVAVELLRKLRARTGVELPATVTFDHPSPHR
LATYVREALPLDAPLAPADATTARSSSAHAAIDEPIAIVGLGLRLPGGVEDREGLWSF
LEQGRDGVGPIPKDRWDADAIYDPDPEAKGKSYVREAALLDRVDLFDAAFFGISPREA
KQVDPQHRLLLETAWQALEEATIVPASLKDTSTGVFVGIGLSDYALMQGSAEETDAYA
AMGTLSSFAAGRLAFTLGLQGPALSIDTACSSSLVALHLACQSLRKGECDLALAGGVQ
VMTLAEPFVLLSRTRALAADGRSKTFSAHADGYGRGEGVVVLALERLRDAEARGRRVL
AVVRGTAVNHDGASSGITAPNGTSQQKVLRAALGDAGLAPADVDVVECHGTGTALGDP
IEVQALAAVYGEGRAAERPLLLGAVKSNIGHLEAASGLAGVAKIVTALEHGALPATIH
SRPRNPHVDWDALHVRVVDAAELWARRDEETPRRAGVSAFGLSGTNAHAIIEEAPRAR
RGAAATDEPPPAVPVVLSAKSERALRAQAERLRAHVTGRPGDALVDVAASLATTRSHF
EHRAAIVARDLGALEDALAGLAEGRSPAGTVVSRTTGGKLALLFTGQGGQRAAMGRGL
YGAFSVFRDAFDAACSYLDRELEHPLRDVVFAPRGSELAALLDETLYTQTSLFALEVA
LFRLVEAWGVTPHVLLGHSIGELVAAHVAGVLSLEDACTLVAARARLMHALPRRDAAM
VAVHACESEVLDALDAMAGLDGRASIAAVNAPSSTVVAGDREAVLAVAARFEALGRKT
SRLRVSHAFHSHHLDPMLEAFRRVAEGVRFEPPRIPIVSNVTGARASNDDLCSADYWV
RHARGTVRFADGLEALRADGVSTFLELGPHAVLCALGHETLESEAVAFLPALRDGRDD
VDTLTAALGALHARGHHVDWDAFFAPFNAGRVALPTYAFQRERFWLDMPKAKSSRAAA
ADSAGRYPLAGTRLDLPDGSAVHTLDIGPGAQPYLADHSVYGRIVVPGAFHLAVLLAV
AESHWPDSPLELRDVEFVRALTFEEPSETVAVRVHLTPAGGDGKGFHATVATQTEGAW
TTHATATLGAAPPSALARHPRAAHEPLAGDEPSRAVARLMASFEAFNVEWGPKWRWLR
QTTHAGDRTGVGYLAAPDGVPTDDAPLPGGLVDNAFAVVGSAVTGLPAHDNRSTSNVP
PLPFSIERLVWYGRAETPCWADFVMRGEITPDADRYIGDITLWSREGASSEGASREGT
PIATIEGLAFRRAPADKFLPELPARNLYALTWPEQPAASRPPRGRWALLGEDSLGIAT
GIAADAYPDVDALRDALGRGAPCPDTVAVCLATAHGPSDTDETELVSSVHTAAARALA
LLNAWLSDERLASSTLVLLTQRAVATCAGEDVLDLASAALWGLVRSAQAENRDRAILL
VDLDGSDASRRALPYAVGGALDDAEPQLALRDGRRLAPRLARVAAPTAATANAATSTS
ATPRALDPEGTVLVTGGTGTLGALVARRLVERHGVKHLVLLSRRGPDAPGATDLAAEL
QARGASVVLAAADAAEPAALARVLLAIPHDRPLTAVVHAAGTLDDGVLSSMTPARLSA
VLRAKVDAAVNLDEQTRHSPLRAFVLFSSLSGVLGSPAQSNYAAANAFLDALAHRRAA
RGLPAVALDWGYWAQTSGLTKHLEQADLRRIARGGLRPLASDEALALFDLALAQDAPC
LVPAAFDADALRKSGDALSALLRGLVAGPRARAQRASGAGAASALKLRLASLAPLERH
RMLLDVVRADVATVLGVASPLDIEPGRPLQELGLDSLMAIELRNRLAAATGLRLSATL
LFDHPTPEALAKLISGKLLHDEAAPPSTSAAEIDRLEGALASAYANEAIRDELTVRMK
AILSKWMAAREPSPDDGLASRLDTANDDELFRMIDHVRAEAAT"

gene 22592. .33133
/gene="spiI"

```
  Query Match             10.6%;  Score 95.6;  DB 1;  Length 48000;
  Best Local Similarity   47.8%;  Pred. No. 16;
  Matches  374;  Conservative    0;  Mismatches  399;  Indels    9;  Gaps    3;

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
              ||| ||||| | | | | ||||| ||    | || |   | | ||||| | | |
Db       6994 GCGCGCGTCGTCCGGTGGCTCGCGCAGCGCGTCGAGCGCCGATCGCGCCTCCGCGAGCTC 6935

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              || || ||        | ||| ||  ||||   | |||    |||    |||||  |||
Db       6934 GTCGAGGAGCGGGCGAGGACGCGCCGAGGCGAACGACGG---GGCGAAGCGCGCCCAGTC 6878

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
              | | || || |||| ||| |  | |   ||       || | |||   ||| | ||    |
Db       6877 GACGTCAGCGACGGCGAGGCACGTCTCGGCGTGATCGAGCGCGTGCGCGAGGGCCGCGAC 6818

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
              |      ||  |  ||       |||||  | |||     | |||  | ||  | ||
Db       6817 GGCGAGCTCCGGATCCATCGGCGCGACGCCGCGATGCTTCAAGCGCCGCTCGGCGTCGGT 6758

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              | | |   | |||| |  | |   || |    |||| | |      ||  |||  ||
Db       6757 CGCGAGCATCCCGCCGCCCGCCCACGCGCCCCACGCGATCGAGGTCGCCGCGAGGCCGTC 6698

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
              |||| | | |  | ||  ||| ||| |  | | | |   ||   |    | | |
Db       6697 CGCGCGACGCGCCTCGGCGAGCGCGTCGAGGAACGCGTTCGCCGCGGCGTACGCGCCTTG 6638

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              | | |  ||||  | |   | || ||     |   |   ||     |  |||| | |
Db       6637 CCGCCCACCGCCCCACACACCCGCTCCCGACGAGAAGAGGACGAACGCGTCGAGCGGGCG 6578

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
              || |  || || | |  | | ||   || |   | |        | |      || ||
Db       6577 CTGCGCGAAGACGTCGTGGAGGTGCTGCGCGCCGCGGGCCTTCGCGGCGATCGCGTCGGC 6518

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
              || ||  |||||||     || || ||  | ||     ||  |||  ||  | | | |
Db       6517 GAGCTCCTCGACACGCATGGCGGCGACCGGCGCCTCGTGCATCGCACCGCCCGCGTGGAA 6458

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
                  | ||| | |       ||  ||| || ||   ||        | | | |   |
Db       6457 CACGGCGCGCACCGTCGCGCCGCCGGCCTCGACGTCGCGCACGAGCGCCGCGACCGCCTC 6398

Qy        726 ATGACGCGCGCGCTCG-ACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGA-----CCCC 779
              | ||   |||  ||| |||| ||| ||||    ||||   ||  || |||      || |
Db       6397 ACGATCGGCGACGTCGCACGCAGCCAGCGTCACGCGCGCGCCGAGCGCCGAGAGCTCCTC 6338

Qy        780 GCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATC 839
              | ||||| | ||| || |  |  | | ||| | | |   ||||  |||||  ||| ||
Db       6337 GGCGAGCGCGGCCGCGCCGGGGGCCGCGGCTCCGCGGCGGCTCGCGAGCACGATGTGCTC 6278

Qy        840 GGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCG 899
              |||   ||    | |  ||       | | || |   | |       | | || ||||
Db       6277 GGCGCCGATCCGCGCGAGCCAGCGGGCAACGTGCGCGCCGAGCGCGCCGGTCCCGCCCGT 6218

Qy        900 GA 901
              ||
Db       6217 GA 6216


RESULT 7
AR217867
LOCUS       AR217867               1926 bp    DNA     linear   PAT 25-SEP-2002
DEFINITION  Sequence 4 from patent US 6417002.
ACCESSION   AR217867
VERSION     AR217867.1  GI:23317761
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1926)
  AUTHORS   Horlick,R.A. and Chelsky,D.
  TITLE     Method for maintenance and selection of episomes
  JOURNAL   Patent: US 6417002-A 4 09-JUL-2002;
            Pharmacopeia, Inc.; Cranbury, NJ
FEATURES             Location/Qualifiers
     source          1. .1926
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             10.6%;  Score 95.4;  DB 6;  Length 1926;
  Best Local Similarity   46.3%;  Pred. No. 59;
  Matches  352;  Conservative    0;  Mismatches  406;  Indels    3;  Gaps    1;

Qy        112 GCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACC 171
              |||| |   |  ||  |||||  | | | || ||  ||   | | |   |    ||
Db        258 GCCACCTTGTCCTCGTCCTCGTCCTCGCCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 317

Qy        172 TCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCG 231
              |||||| |     | |||     | | ||  ||  ||| |  || |  ||   ||| ||
Db        318 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCG 377

Qy        232 CGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCG 291
              | | || |||| || ||  || |   |  | |       | || |       || ||
Db        378 TCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 437

Qy        292 TCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGG 351
              ||  |||||    |   |||  |  | ||| | |  || |        || |    | |
Db        438 TCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCG 497

Qy        352 TGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCG 411
              |  || || |||||| || |||| | | |  |||  |   |  ||     |   | |  ||
Db        498 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCG 557

Qy        412 CGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGA 471
              |  |    |  ||  | |    | | || ||  |  |||| | ||  |||||  |
Db        558 TCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCC--- 614

Qy        472 TCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGG 531
              |||  |    |   ||| || || |||| ||   | | |||  | |  |  |  |
Db        615 TCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC 674

Qy        532 AGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGA 591
              | | |  |||  | |||        |   |  |||||  || | |||| |   |
Db        675 TCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCG 734

Qy        592 CCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCC 651
              || | |  ||  |    ||  || |      | ||  || || || |   |   ||| ||
Db        735 TCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 794

Qy        652 GCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCG 711
              |||  |||| ||          || || | ||  |    |   |||  |  |     | ||
Db        795 CCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC 854

Qy        712 CGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCG 771
              |    | | | | |  |    | || |   | || |    |||  |  ||   |||
Db        855 TCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCG 914

Qy        772 ACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACG 831
              |  ||||| |   | |  |||||||| | |   | |   || |      |    |
Db        915 TCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 974

Qy        832 ACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAG 872
              |    || |    || ||| ||| | | |||||  | ||
Db        975 TCCACCTCCGGCCCCAGCTCCTCCGTCACCTCCGGCCCCAG 1015


RESULT 8
AR083152
LOCUS       AR083152               1931 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 2 from patent US 5976807.
ACCESSION   AR083152
VERSION     AR083152.1  GI:10009942
```

```
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1931)
  AUTHORS   Horlick,R.A., Damaj,B.B. and Robbins,A.K.
  TITLE     Eukaryotic cells stably expressing genes from multiple transfected
            episomes
  JOURNAL   Patent: US 5976807-A 2 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1931
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             10.6%;  Score 95.4;  DB 6;  Length 1931;
  Best Local Similarity   46.3%;  Pred. No. 59;
  Matches  352;  Conservative    0;  Mismatches  406;  Indels    3;  Gaps    1;

Qy        112 GCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACC 171
              |||| |    |  ||   |||||  | | | || ||  ||    || |  |    ||
Db        258 GCCACCTTGTCCTCGTCCTCGTCCTCGCCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 317

Qy        172 TCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCG 231
              |||||| |      | |||      | | ||   ||  ||| |  || |  ||   ||| ||
Db        318 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCG 377

Qy        232 CGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCG 291
                | | || |||| || ||  || |   |  | |       | || |       || ||
Db        378 TCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 437

Qy        292 TCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGG 351
              ||  |||||    |   |||  |   | ||| | |  || |         || |    | |
Db        438 TCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCG 497

Qy        352 TGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCG 411
              |  || || |||||| || |||| | | |  ||||  |  |  ||    |   | |  ||
Db        498 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCG 557

Qy        412 CGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGA 471
                |  |   |  ||  | |    | | || ||  |  ||||  | ||  |||||  |
Db        558 TCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCC--- 614

Qy        472 TCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGG 531
              ||| |      |    ||| || || |||| ||   | | |||  | |   |  | |
Db        615 TCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC 674

Qy        532 AGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGA 591
                | |  |   |||  |  |||      |   |  |||||  || | |||| |   |
Db        675 TCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCG 734

Qy        592 CCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCC 651
              || | |  ||  |    ||  || | |    | ||  || || || |   |   ||| ||
Db        735 TCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 794

Qy        652 GCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCG 711
              ||| |||| ||          || || | || |    |  ||| |  |    | ||
Db        795 CCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC 854

Qy        712 CGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCG 771
                |   | |  | |  |   |  | || |   | ||| |    ||| |  ||   |||
Db        855 TCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCG 914

Qy        772 ACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACG 831
              |  ||||| |   | |  ||||||||| | |   | |   || |       |     |
Db        915 TCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 974

Qy        832 ACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAG 872
              |    || |     || ||| ||| | | |||||  | ||
Db        975 TCCACCTCCGGCCCCAGCTCCTCCGTCACCTCCGGCCCCAG 1015


RESULT 9
```

```
SCO939123
LOCUS       SCO939123            300100 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 20/29.
ACCESSION   AL939123 AL031184 AL035569 AL079356 AL137242 AL138978 AL157916
            AL359152 AL359214 AL391588 AL391763 AL589707 AL645882
VERSION     AL939123.1  GI:24430032
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
   PUBMED   12000953
REFERENCE   2  (bases 1 to 300100)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 30, 2002 this sequence version replaced
            gi:20520687, gi:20520766, gi:20520826, gi:20520797, gi:20520798,
            gi:20520697, gi:20520859, gi:20520743, gi:20520862.
FEATURES             Location/Qualifiers
     source          1. .300100
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     RBS             53. .56
     gene            66. .1151
                     /gene="SCO5222"
                     /note="synonym: SC7E4.19"
     CDS             66. .1151
                     /gene="SCO5222"
                     /note="SC7E4.19, possible lyase, len: 361 aa; similar to
                     SW:PTLS_STRSQ (EMBL:U05213) Streptomyces sp. pentalenene
                     synthase (EC 4.6.1.5), 336 aa; fasta scores: opt: 254
                     z-score: 317.5 E(): 3.3e-10; 23.8% identity in 311 aa
                     overlap. Contains match to Prosite entry PS00904 Protein
                     prenyltransferases alpha subunit repeat signature and a
                     TTA leucine codon, possible target for bldA regulation"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lyase"
                     /protein_id="CAB94607.1"
                     /db_xref="GI:8546888"
                     /db_xref="GOA:Q9K499"
                     /db_xref="InterPro:IPR002088"
                     /db_xref="InterPro:IPR005630"
                     /db_xref="UniProt/TrEMBL:Q9K499"
                     /translation="MHAFPHGTTATPTAIAVPPSLRLPVIEAAFPRQLHPYWPKLQET
                     TRTWLLEKRLMPADKVEEYADGLCYTDLMAGYYLGAPDEVLQAIADYSAWFFVWDDRH
                     DRDIVHGRAGAWRRLRGLLHTALDSPGDHLHHEDTLVAGFADSVRRLYAFLPATWNAR
                     FARHFHTVIEAYDREFHNRTRGIVPGVEEYLELRRLTFAHWIWTDLLEPSSGCELPDA
                     VRKHPAYRRAALLSQEFAAWYNDLCSLPKEIAGDEVHNLGISLITHHSLTLEEAIGEV
                     RRRVEECITEFLAVERDALRFADELADGTVRGKELSGAVRANVGNMRNWFSSVYWFHH
                     ESGRYMVDSWDDRSTPPYVNNEAAGEK"
     misc_feature    327. .356
                     /gene="SCO5222"
                     /note="PS00904 Protein prenyltransferases alpha subunit
                     repeat signature"
```

gene 948. .950
/gene="none"
misc_feature 948. .950
/gene="none"
/note="TTA leucine codon. Possible target for bldA regulation"
gene 1148. .2533
/gene="SCO5223"
/note="synonym: SC7E4.20"
CDS 1148. .2533
/gene="SCO5223"
/note="SC7E4.20, probable cytochrome P450, len: 461 aa; similar to SW:CP51_HUMAN (EMBL:U23942) Homo sapiens cytochrome P450 51 (EC 1.14.14.1) Cyp51, 503 aa; fasta scores: opt: 532 z-score: 587.2 E(): 3.1e-25; 25.7% identity in 447 aa overlap. Contains Pfam match to entry PF00067 p450, Cytochrome P450"
/codon_start=1
/transl_table=11
/product="putative cytochrome P450"
/protein_id="CAB94608.1"
/db_xref="GI:8546889"
/db_xref="GOA:Q9K498"
/db_xref="InterPro:IPR001128"
/db_xref="InterPro:IPR002401"
/db_xref="UniProt/TrEMBL:Q9K498"
/translation="MTVESVNPETRAPAAPGAPELRLPPPVAGGGVPLLGHGWRLARDP
LAFMSQLRDHGDVVRIKLGPKTVYAVTNPELTGALALNPDYHIAGPLWESLEGLLGKE
GVATANGPLHRRQRRTIQPAFRLDAIPAYGPIMEEEAHALTERWQPGKTVDATSESFR
VAVRVAARCLLRGQYMDERAERLCVALATVFRGMYRRMVVPLGPLYRLPLPANRRFND
ALADLHLLVDEIIAERRASGQKPDDLLTALLEAKDDNGDPIGEQEIHDQVVAILTPGS
ETIASTIMWLLQALADHPEHADRIRDEVEAVTGGRPVAFEDVRKLRHTGNVIVEAMRL
RPAVWVLTRRAVAESELGGYRIPAGADIIYSPYAIQRDPKSYDDNLEFDPDRWLPERA
ANVPKYAMKPFSAGKRKCPSDHFSMAQLTLITAALATKYRFEQVAGSNDAVRVGITLR
PHDLLVRPVAR"
misc_feature 1217. .2428
/gene="SCO5223"
/note="Pfam match to entry PF00067 p450, Cytochrome P450, score 221.80, E-value 1e-62"
gene complement(2547. .3515)
/gene="SCO5224"
/note="synonym: SC7E4.21c"
CDS complement(2547. .3515)
/gene="SCO5224"
/note="SC7E4.21c, probable araC-family transcriptional regulator, len: 322 aa; similar to TR:Q9S166 (EMBL:AB023785) Streptomyces griseus transcriptional activator for strR, AdpA, 405 aa; fasta scores: opt: 1005 z-score: 1177.5 E(): 0; 49.2% identity in 325 aa overlapand to TR:CAB87229 (EMBL:AL163641) Streptomyces coelicolor araC-family transcriptional regulator AdpA, 398 aa; fasta scores: opt: 990 z-score: 1160.1 E(): 0; 48.9% identity in 313 aa overlap. Contains Pfam match to entry PF00165 HTH_AraC, Bacterial regulatory helix-turn-helix proteins, araC family and match to Prosite entry PS00041 Bacterial regulatory proteins, araC family signature. Contains also a possible helix-turn-helix motif at residues 230. .251 (+3.72 SD)"
/codon_start=1
/transl_table=11
/product="putative araC-family transcriptional regulator"
/protein_id="CAB94609.1"
/db_xref="GI:8546890"
/db_xref="GOA:Q9K497"
/db_xref="InterPro:IPR000005"
/db_xref="InterPro:IPR002818"
/db_xref="UniProt/TrEMBL:Q9K497"
/translation="MLRNVAAVLLDGAHPFELGVVCEVFGIDRSDEGLPVYDFAVVSA
EGPTLGTHVGGLTVSTPYGLERLEEADLIAVPAGSDFVRREYPPDLLDALRRATDRGT
RVLSVCSGVFVLGAAGLLDGRQCAVHWHHAAELARQHPRARVAPDVLYVDEDPVVTSA
GTAAGIDACLHIVRKEQGPEVANRIARRMVVPPHRDGGQAQYIERPLPRSSCDTVGEV
LAWMEQHLDEEVTVEQLAVRAHMSPRTFARRFQQETGTTPYRWILRQRVLLAQRLLEA
TDETMDTIAWRAGFGTAAALRHQFTRALDTTPHAYRRTFRGPEAAA"

misc_feature complement(2571. .2831)
/gene="SCO5224"
/note="Pfam match to entry PF00165 HTH_AraC, Bacterial regulatory helix-turn-helix proteins, araC family, score 79.70, E-value 6e-20"
misc_feature complement(2592. .2720)
/gene="SCO5224"
/note="PS00041 Bacterial regulatory proteins, araC family signature"
misc_feature complement(3557. .7543)
/note="previously sequenced DNA fragment. EMBL:AJ276618 Streptomyces coelicolor A3(2) nrdX gene, nrdL gene and nrdM gene"
gene complement(3787. .4818)
/gene="SCO5225"
/note="synonyms: nrdM, SC7E4.22c"
CDS complement(3787. .4818)
/gene="SCO5225"
/note="SC7E4.22c, nrdM, ribonucleotide-diphosphate reductase small chain, len: 343 aa; identical to previously sequenced TR:CAB82486 (EMBL:AJ276618) Streptomyces coelicolor A3(2) ribonucleotide-diphosphate reductase small chain (EC 1.17.4.1) NrdM, 343 aa. Contains Pfam match to entry PF00268 ribonuc_red, Ribonucleotide reductases and a possible hydrophobic membrane spanning region"
/codon_start=1
/transl_table=11
/product="ribonucleotide-diphosphate reductase small chain"
/protein_id="CAB94610.1"
/db_xref="GI:8546891"
/db_xref="GOA:Q9K3C2"
/db_xref="InterPro:IPR000358"
/db_xref="UniProt/TrEMBL:Q9K3C2"
/translation="MTTETPAHPTTKNLLDPGFELTLRPMRYPDFYERYRDAIKNTWH
VEEVDLHSDVADLAKLSPMEQHLIGRLVAFFATGDSIVANNLVLTLYKHINSPEARLY
LSRQLFEEAVHVQFYLTLLDTYLPDPEDRAAAFAAVENIPSIREKAEFCFRWMDSVES
IDRLETKADRRRFLLNLICFAACIEGLFFYGAFAYVYWFRSRGLLHGLATGTNWVFRD
ETMHMSFAFDVVDTVRKEEPELFDDQLREQVTDMLREAVEAELQFARDLCGDGLPGMN
TDSMRQYLECVADQRLTRLGFAPVYGSENPFSFMELQGVQELTNFFERRPSAYQVAVE
GTVDLSEDF"
misc_feature complement(3832. .4692)
/gene="SCO5225"
/note="Pfam match to entry PF00268 ribonuc_red, Ribonucleotide reductases, score 222.30, E-value 1e-75"
gene complement(4818. .7193)
/gene="SCO5226"
/note="synonyms: nrdL, SC7E4.23c"
CDS complement(4818. .7193)
/gene="SCO5226"
/note="SC7E4.23c, nrdL, ribonucleotide-diphosphate reductase large chain, len: 791 aa; identical to previously sequenced TR:CAB82485 (EMBL:AJ276618) Streptomyces coelicolor A3(2) ribonucleotide-diphosphate

Query Match 10.5%; Score 94.8; DB 1; Length 300100;
Best Local Similarity 46.7%; Pred. No. 9;
Matches 340; Conservative 0; Mismatches 382; Indels 6; Gaps 1;

```
Qy      122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
            || | | | |  |||| |  | ||     | || |    ||  |||||||||| ||
Db   231955 CGGCCCCGGCACCGGCAAGACCGCCGTCGCCCTGCACCGCGCCGCCTACCTCCTCTACGA 232014

Qy      182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
            |  || |    ||||  | |||    |  |  |  || | |   | | ||     ||
Db   232015 GCACCGCGAACTGCTCGCCAAGCGCGCCGTCCTCATCGTCGGCCCCAACCCGGCCTTCCT 232074

Qy      242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
            || |  ||||||| | | |  ||||| ||   || |    |   | || ||| |  |
Db   232075 CGGCTACATCGGCGAGGTGCTGCCCTCGCTCGGCGAGACCGGCGTGCTCCTCGCCACCGT 232134

Qy      302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
```

```
Db     232135 CGGCGAGCTGTTCCCCGGCGTGAAGGCCACGGCGACGGACACCCCCGAGGCGGCGGCCGT 232194

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421

Db     232195 GAAGGGCCGGGCCGAGATGGCCGACGTGCTCGCGGAGGCGGTCCGCGCCCGGCAGGCACT 232254

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481

Db     232255 GCCCGACCCGGTGATCGCGATCGAGCACGACCGCGACATCCTC------ATGCTCGACGA 232308

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541

Db     232309 CGACCTGGTCAACGTCGCCCGCGAGCGCACCCGCGCCGCCAAGCTGCCGCACAACGCGGC 232368

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601

Db     232369 CCGCGAACACTTCGAGGGACACATCCTCAACACCCTCACCGACATGGTCGCCGAGCGCAT 232428

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661

Db     232429 CGGCACCGACCCCTACGACGGCAGCAACCTCCTCGACCCCAGCGACATCACCCAGATCCG 232488

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721

Db     232489 CGACGAACTCGCCGAGAACCCCGAGGTCTGGTCCGCCGTCGACCGGCTGTGGCCCCGGCT 232548

Qy        722 CACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGC 781

Db     232549 CACGCCGCAGCGCCTGGTCGCCGACCTCCTCGCCGCCCCGGAGTCCTTCCTCCCCGCCGA 232608

Qy        782 CGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGG 841

Db     232609 GGACGCCGCCGCCGTCCGCCGCCCGGTCACCCGGCGCTGGACCGTCGCCGACGTGCCCCT 232668

Qy        842 CAGGGAGG 849

Db     232669 GCTGGACG 232676


RESULT 10
AX470092
LOCUS       AX470092                1869 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 2 from Patent WO02053750.
ACCESSION   AX470092
VERSION     AX470092.1  GI:22205324
KEYWORDS    .
SOURCE      Streptomyces coelicolor
  ORGANISM  Streptomyces coelicolor
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Chouayekh,H. and Virolle,M.N.
  TITLE     Micro-organism strains for implementing a method for overproduction
            of metabolites, in particular with antibiotic or enzymatic activity
  JOURNAL   Patent: WO 02053750-A 2 11-JUL-2002;
            CENTRE NAT RECH SCIENT (FR)
FEATURES             Location/Qualifiers
     source          1. .1869
                     /organism="Streptomyces coelicolor"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1902"
ORIGIN

  Query Match           10.1%;  Score 91.4;  DB 6;  Length 1869;
  Best Local Similarity   46.8%;  Pred. No. 1.2e+02;
  Matches  324;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172

Db        666 CCTCACCCCCCTCGCGGTCGACCCCGCTCACCCCTTCCCGTACATCTCGGGCCTCTCCCT 725

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232

Db        726 CAACCTCGCGGTCGTCGTCCGCAACCCCGTCACCGGCCACCGCCACTTCGCCCGCGTCAA 785

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292

Db        786 GGTCCCCCCACTGCTCTCCCGCTTCCTGGAGGCCTCCCCGGGCCGCTACGTCCCCGTCGA 845

Qy        293 CGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGT 352

Db        846 GGACGTCATCGCCGCCCACCTGGAGGAGCTGTTCCCGGGCATGGAGGTCCTGGAGCACCA 905

Qy        353 GTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGC 412

Db        906 CACCTTCCGCCTCACCCGCAACGAGGACCTGGAGGTCGAGGAGGACGACGCGGAGAACCT 965

Qy        413 GCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGAT 472

Db        966 GCTCCAGGCCCTGGAGAAGGAGCTGATGCGGCGCCGCCTGGGCCCGCCCGTCCGACTGGA 1025

Qy        473 CCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGA 532

Db       1026 GGTCGAGGAGTCCGTCGACCGCGAGGTCCTGGACCTCCTCGTACGCGAGCTGAAGATCGG 1085

Qy        533 GGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGAC 592

Db       1086 CGAGGCGGAGGTCTACCCGCTCCCCGGCCCCCTGGACCTCACCGGCCTCTTCCGCATCCA 1145

Qy        593 CCACATGATCAG---CAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCG 649

Db       1146 CAGCATCGACCGGCCCGAGCTGAAGTACCCGAAGTTCGTCGGCCGGCACCCACCGCGACCT 1205

Qy        650 CCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCT 709

Db       1206 GGCCGAGGTCGAGTCCGCGTCCGCGCCGGACATCTTCGCCGCCCTGCGCACCAAGGACGT 1265

Qy        710 CGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGT 769

Db       1266 CCTGCTCCACCACCCGTACGACTCCTTCTCCACCTCCGTGCAGGCGTTCCTGGAGCAGGC 1325

Qy        770 CGACGACCCCGCCGAGCTCCGCCTCGTCCTGCA 802

Db       1326 CGCCGCCGACCCGGACGTCCTCGCGATCAAGCA 1358


RESULT 11
AX470093
LOCUS       AX470093                2126 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 3 from Patent WO02053750.
ACCESSION   AX470093
VERSION     AX470093.1  GI:22205325
KEYWORDS    .
SOURCE      Streptomyces coelicolor
  ORGANISM  Streptomyces coelicolor
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Chouayekh,H. and Virolle,M.N.
  TITLE     Micro-organism strains for implementing a method for overproduction
            of metabolites, in particular with antibiotic or enzymatic activity
  JOURNAL   Patent: WO 02053750-A 3 11-JUL-2002;
            CENTRE NAT RECH SCIENT (FR)
FEATURES             Location/Qualifiers
     source          1. .2126
                     /organism="Streptomyces coelicolor"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1902"
ORIGIN

  Query Match           10.1%;  Score 91.4;  DB 6;  Length 2126;
  Best Local Similarity   46.8%;  Pred. No. 1.1e+02;
  Matches  324;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172
```

```
Db        923 CCTCACCCCCCTCGCGGTCGACCCCGCTCACCCCTTCCCGTACATCTCGGGCCTCTCCCT 982

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232

Db        983 CAACCTCGCGGTCGTCGTCCGCAACCCCGTCACCGGCCACCGCCACTTCGCCCGCGTCAA 1042

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292

Db       1043 GGTCCCCCCACTGCTCTCCCGCTTCCTGGAGGCCTCCCCGGGCCGCTACGTCCCCGTCGA 1102

Qy        293 CGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGT 352

Db       1103 GGACGTCATCGCCGCCCACCTGGAGGAGCTGTTCCCGGGCATGGAGGTCCTGGAGCACCA 1162

Qy        353 GTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGC 412

Db       1163 CACCTTCCGCCTCACCCGCAACGAGGACCTGGAGGTCGAGGAGGACGACGCGGAGAACCT 1222

Qy        413 GCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGAT 472

Db       1223 GCTCCAGGCCCTGGAGAAGGAGCTGATGCGGCGCCGCCTGGGCCCGCCCGTCCGACTGGA 1282

Qy        473 CCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGA 532

Db       1283 GGTCGAGGAGTCCGTCGACCGCGAGGTCCTGGACCTCCTCGTACGCGAGCTGAAGATCGG 1342

Qy        533 GGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGAC 592

Db       1343 CGAGGCGGAGGTCTACCCGCTCCCCGGCCCCCTGGACCTCACCGGCCTCTTCCGCATCCA 1402

Qy        593 CCACATGATCAG---CAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCG 649

Db       1403 CAGCATCGACCGGCCCGAGCTGAAGTACCCGAAGTTCGTCGCCGGCACCCACCGCGACCT 1462

Qy        650 CCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCT 709

Db       1463 GGCCGAGGTCGAGTCCGCGTCCGCGCCGGACATCTTCGCCGCCCTGCGCACCAAGGACGT 1522

Qy        710 CGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGT 769

Db       1523 CCTGCTCCACCACCCGTACGACTCCTTCTCCACCTCCGTGCAGGCGTTCCTGGAGCAGGC 1582

Qy        770 CGACGACCCCGCCGAGCTCCGCCTCGTCCTGCA 802

Db       1583 CGCCGCCGACCCGGACGTCCTCGCGATCAAGCA 1615


RESULT 12
AX470091
LOCUS       AX470091               3780 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 1 from Patent WO02053750.
ACCESSION   AX470091
VERSION     AX470091.1  GI:22205323
KEYWORDS    .
SOURCE      Streptomyces coelicolor
  ORGANISM  Streptomyces coelicolor
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Chouayekh,H. and Virolle,M.N.
  TITLE     Micro-organism strains for implementing a method for overproduction
            of metabolites, in particular with antibiotic or enzymatic activity
  JOURNAL   Patent: WO 02053750-A 1 11-JUL-2002;
            CENTRE NAT RECH SCIENT (FR)
FEATURES             Location/Qualifiers
     source          1. .3780
                     /organism="Streptomyces coelicolor"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1902"
ORIGIN

  Query Match             10.1%;  Score 91.4;  DB 6;  Length 3780;
  Best Local Similarity   46.8%;  Pred. No. 89;
  Matches  324;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172

Db       1666 CCTCACCCCCCTCGCGGTCGACCCCGCTCACCCCTTCCCGTACATCTCGGGCCTCTCCCT 1725

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232

Db       1726 CAACCTCGCGGTCGTCGTCCGCAACCCCGTCACCGGCCACCGCCACTTCGCCCGCGTCAA 1785

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292

Db       1786 GGTCCCCCCACTGCTCTCCCGCTTCCTGGAGGCCTCCCCGGGCCGCTACGTCCCCGTCGA 1845

Qy        293 CGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGT 352

Db       1846 GGACGTCATCGCCGCCCACCTGGAGGAGCTGTTCCCGGGCATGGAGGTCCTGGAGCACCA 1905

Qy        353 GTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGC 412

Db       1906 CACCTTCCGCCTCACCCGCAACGAGGACCTGGAGGTCGAGGAGGACGACGCGGAGAACCT 1965

Qy        413 GCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGAT 472

Db       1966 GCTCCAGGCCCTGGAGAAGGAGCTGATGCGGCGCCGCCTGGGCCCGCCCGTCCGACTGGA 2025

Qy        473 CCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGA 532

Db       2026 GGTCGAGGAGTCCGTCGACCGCGAGGTCCTGGACCTCCTCGTACGCGAGCTGAAGATCGG 2085

Qy        533 GGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGAC 592

Db       2086 CGAGGCGGAGGTCTACCCGCTCCCCGGCCCCCTGGACCTCACCGGCCTCTTCCGCATCCA 2145

Qy        593 CCACATGATCAG---CAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCG 649

Db       2146 CAGCATCGACCGGCCCGAGCTGAAGTACCCGAAGTTCGTCGCCGGCACCCACCGCGACCT 2205

Qy        650 CCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCT 709

Db       2206 GGCCGAGGTCGAGTCCGCGTCCGCGCCGGACATCTTCGCCGCCCTGCGCACCAAGGACGT 2265

Qy        710 CGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGT 769

Db       2266 CCTGCTCCACCACCCGTACGACTCCTTCTCCACCTCCGTGCAGGCGTTCCTGGAGCAGGC 2325

Qy        770 CGACGACCCCGCCGAGCTCCGCCTCGTCCTGCA 802

Db       2326 CGCCGCCGACCCGGACGTCCTCGCGATCAAGCA 2358


RESULT 13
BA000030_47/c
WPCOMMENT
  Sequence split into 91 fragments  LOCUS BA000030 Accession BA000030
     Fragment Name      Begin     End
     BA000030_00            1    110000
     BA000030_01       100001    210000
     BA000030_02       200001    310000
     BA000030_03       300001    410000
     BA000030_04       400001    510000
     BA000030_05       500001    610000
     BA000030_06       600001    710000
     BA000030_07       700001    810000
     BA000030_08       800001    910000
     BA000030_09       900001   1010000
     BA000030_10      1000001   1110000
     BA000030_11      1100001   1210000
     BA000030_12      1200001   1310000
     BA000030_13      1300001   1410000
     BA000030_14      1400001   1510000
     BA000030_15      1500001   1610000
```

BA000030_16	1600001	1710000
BA000030_17	1700001	1810000
BA000030_18	1800001	1910000
BA000030_19	1900001	2010000
BA000030_20	2000001	2110000
BA000030_21	2100001	2210000
BA000030_22	2200001	2310000
BA000030_23	2300001	2410000
BA000030_24	2400001	2510000
BA000030_25	2500001	2610000
BA000030_26	2600001	2710000
BA000030_27	2700001	2810000
BA000030_28	2800001	2910000
BA000030_29	2900001	3010000
BA000030_30	3000001	3110000
BA000030_31	3100001	3210000
BA000030_32	3200001	3310000
BA000030_33	3300001	3410000
BA000030_34	3400001	3510000
BA000030_35	3500001	3610000
BA000030_36	3600001	3710000
BA000030_37	3700001	3810000
BA000030_38	3800001	3910000
BA000030_39	3900001	4010000
BA000030_40	4000001	4110000
BA000030_41	4100001	4210000
BA000030_42	4200001	4310000
BA000030_43	4300001	4410000
BA000030_44	4400001	4510000
BA000030_45	4500001	4610000
BA000030_46	4600001	4710000
BA000030_47	4700001	4810000
BA000030_48	4800001	4910000
BA000030_49	4900001	5010000
BA000030_50	5000001	5110000
BA000030_51	5100001	5210000
BA000030_52	5200001	5310000
BA000030_53	5300001	5410000
BA000030_54	5400001	5510000
BA000030_55	5500001	5610000
BA000030_56	5600001	5710000
BA000030_57	5700001	5810000
BA000030_58	5800001	5910000
BA000030_59	5900001	6010000
BA000030_60	6000001	6110000
BA000030_61	6100001	6210000
BA000030_62	6200001	6310000
BA000030_63	6300001	6410000
BA000030_64	6400001	6510000
BA000030_65	6500001	6610000
BA000030_66	6600001	6710000
BA000030_67	6700001	6810000
BA000030_68	6800001	6910000
BA000030_69	6900001	7010000
BA000030_70	7000001	7110000
BA000030_71	7100001	7210000
BA000030_72	7200001	7310000
BA000030_73	7300001	7410000
BA000030_74	7400001	7510000
BA000030_75	7500001	7610000
BA000030_76	7600001	7710000
BA000030_77	7700001	7810000
BA000030_78	7800001	7910000
BA000030_79	7900001	8010000
BA000030_80	8000001	8110000
BA000030_81	8100001	8210000
BA000030_82	8200001	8310000
BA000030_83	8300001	8410000
BA000030_84	8400001	8510000
BA000030_85	8500001	8610000
BA000030_86	8600001	8710000
BA000030_87	8700001	8810000
BA000030_88	8800001	8910000
BA000030_89	8900001	9010000
BA000030_90	9000001	9025608

Continuation (48 of 91) of BA000030 from base 4700001 (BA000030 Streptomyces avermitilis

Query Match 10.1%; Score 91.4; DB 1; Length 110000;
Best Local Similarity 47.4%; Pred. No. 24;
Matches 420; Conservative 0; Mismatches 446; Indels 21; Gaps 4;

```
Qy        15 GACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGC 74
             | ||||   |   || |  ||| |||| |  ||    |||||   || |  ||  |   |
Db     79893 GGCGACCGTTTCGACGCCCTCCACGAGATGCTCACCGGCGGCGAACTGGACGTCCTGACC 79834

Qy        75 GGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCG------CCAGCCGCACGACGCG 128
             |||||       ||||||| ||   |||    | || |       ||  | | | ||| |
Db     79833 GGCGACTACCTCGCCGAGCTGACCATGCTCATCCTCGGCCGCGACCGCCTGAAGGACCCC 79774

Qy       129 GTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTC 188
             || ||| |    || | |  | |  ||| || | |   |  |||  ||   ||| |
Db     79773 ACCGCCGGTTACGCCCGCACCTTCCTCCGCCAGCTCGAGGGCTGCCTCGGCCTCGCCCAC 79714

Qy       189 CGGGCGCTCATCGAGCCGCTCCC---CGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
               |  | | |   |    || |   ||  |||||||    || |||    | |  || |
Db     79713 GAGCGGGGCGTGCGGATCGTCACCAACGCGGGCGGCCTCAACCCCGCCGGACTCGCCGAC 79654

Qy       246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
             | |||| |  ||  ||      | |   ||  |      | ||||| | ||||  || ||
Db     79653 GCCATCCGGAACCTGGCCGGACGCCTCGGCATCCCCACCCGCGTCGCCCACGTGGAGGGC 79594

Qy       306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
             ||   | |||    | || || | |        || | | ||     ||   |  | ||
Db     79593 GACGACCTCACCGCCGCGCACCCGGGCAGCCTCGCCGCCCACGCCTACCTCGGCGGCTTC 79534

Qy       366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
             | |  | ||||         ||  | || |||   || || | |||   ||||| || |
Db     79533 GGCATCGCCGCGTGTCTGCGCGAGGGCGCCGACATCGTGGTCACCG---GGCGGGTGACC 79477

Qy       426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
             | |||| | |    |  ||  | |  || ||||||| |||   ||   |   ||||  |
Db     79476 GACGCGGCCCTGGTCACCGGGCCCGCCGCCGCCCACTTCGGCTGGCAGCCGGGGGAGTAC 79417

Qy       486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
               ||  || ||   |  ||   ||  |  | || || |  |   | |||| || | |||
Db     79416 GACCGGCTGGCGGGCGCCGTCGTCGCCGGGCACGTGCTGGAGTGCGGGGCGCAGGCCACC 79357

Qy       546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
             |   ||  | | |      | | |  |   ||| ||   ||  |       |  | ||
Db     79356 GGCGGGAACTACGCCTTCTTCCGCGACGCGGACGTCCGCCGCCCCGGGTTCCCGCTCGCG 79297

Qy       606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGC---------TCCGCCGCGTG 656
             |||||  ||      ||  || || |  |  | |||  |           | ||| |||
Db     79296 GAGATCCACGCGGACGGCACCGGCGTCATCACCAAGCACCCCGGCACCGGCGGCCTCGTC 79237

Qy       657 CTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTC 716
               ||||  ||   | | |||   || || | | ||  |  | ||          ||||| |
Db     79236 GACGTCGGTACGGTCACCGCCCAGTTCCTGTACGAGACGGCGGGCGCCCGGTACGCGGGC 79177

Qy       717 TATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGAC 776
               |  ||   | |||      |  ||      | |   |    || | | | |  || |
Db     79176 CCCGACGTCACCGCCCGCCTCGACACCGTACGGCTCACCCAGGACGGCCCCGACCGCGTC 79117

Qy       777 CCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTG 836
             | || ||||  |   | || | |  | | | | ||| |  | | |  ||  | ||  |
Db     79116 CGCGTCGAGGGCGTACGCGGCGAGGCCCCGCCCCCGGCCCTCAAGGTCGGCCGCAACCGG 79057

Qy       837 ATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGA 883
             |||||  |      | ||  | || ||    ||  | | |    |||
Db     79056 CTCGGCGGCTTCCGCAACGAGGTCGTCTTCGTCCTGACCGGTCTCGA 79010
```

RESULT 14
SCO939119/c
LOCUS SCO939119 299050 bp DNA linear BCT 16-APR-2005
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 16/29.

```
ACCESSION   AL939119 AL160331 AL353816 AL356612 AL357432 AL357524 AL358692
            AL358818 AL359988 AL391406 AL391541 AL392149 AL392150 AL392178
            AL645882
VERSION     AL939119.1  GI:24427855
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
   PUBMED   12000953
REFERENCE   2  (bases 1 to 299050)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 30, 2002 this sequence version replaced
            gi:7242748, gi:7636006, gi:8248766, gi:8249971, gi:8388712,
            gi:8439477, gi:8894718, gi:9857143, gi:9886716, gi:10129746,
            gi:11595528.
FEATURES             Location/Qualifiers
     source          1. .299050
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     gene            200. .1727
                     /gene="rrnA"
     rRNA            200. .1727
                     /gene="rrnA"
                     /note="rrnA, 16S ribosomal RNA. Coordinates based on
                     comparison with rrnB from Streptomyces coelicolor
                     (EMBL:AL109848)."
     gene            2004. .5126
                     /gene="rrnA"
     rRNA            2004. .5126
                     /gene="rrnA"
                     /note="rrnA, 23S ribosomal RNA. Coordinates based on
                     comparison with rrnB from Streptomyces coelicolor
                     (EMBL:AL109848)."
     gene            5228. .5348
                     /gene="rrnA"
     rRNA            5228. .5348
                     /gene="rrnA"
                     /note="rrnA, 5S ribosomal RNA. Coordinates based on
                     comparison with rrnB from Streptomyces coelicolor
                     (EMBL:AL109848)."
     gene            complement(5412. .6080)
                     /gene="SCO4123"
                     /note="synonym: SCD72A.09c"
     CDS             complement(5412. .6080)
                     /gene="SCO4123"
                     /note="SCD72A.09c, possible two-component system response
                     regulator, len: 222aa; similar to many eg. TR:O86632
                     (EMBL:AL031155) possible two-component regulator from
                     Streptomyces coelicolor (222 aa) fasta scores; opt: 973,
                     z-score: 1128.5, E(): 0, 69.1% identity in 220 aa overlap.
                     Contains Pfam match to entry PF00072 response_reg,
                     Response regulator receiver domain, Pfam match to entry
                     PF00196 GerE, Bacterial regulatory proteins, luxR family
                     and Prosite match to PS00622 Bacterial regulatory
                     proteins, luxR family signature."
                     /codon_start=1
                     /transl_table=11
                     /product="putative two-component system response
                     regulator"
                     /protein_id="CAB92376.1"
                     /db_xref="GI:8052435"
                     /db_xref="GOA:Q9KYE7"
                     /db_xref="InterPro:IPR000792"
                     /db_xref="InterPro:IPR001789"
                     /db_xref="UniProt/TrEMBL:Q9KYE7"
                     /translation="MTTVLIADDQPLQRFGFRMLLESQDDMTVVGEAANGAEAARMAA
                     ELHPDVVLMDVRMPGLDGIEATRRIVAGGGRSRVLILTTFDLDEYAYSGLRAGASGFL
                     VKDALPEELLSGIRAVATGDAVVAPSLTRRLLDAYVHHLPAAGTRDHTPGSPDPRLSP
                     LTERETEILRVMGQGWTNTEIAERLHLAESTVKTHVGRILAKTGSRDRVQAVILAYET
                     KLVG"
misc_feature         complement(5415. .5609)
                     /gene="SCO4123"
                     /note="Pfam match to entry PF00196 GerE, Bacterial
                     regulatory proteins, luxR family, score 61.00, E-value
                     2.6e-14"
misc_feature         complement(5475. .5558)
                     /gene="SCO4123"
                     /note="PS00622 Bacterial regulatory proteins, luxR family
                     signature"
misc_feature         complement(5736. .6077)
                     /gene="SCO4123"
                     /note="Pfam match to entry PF00072 response_reg, Response
                     regulator receiver domain, score 126.30, E-value 5.7e-34"
gene                 complement(6077. .7447)
                     /gene="SCO4124"
                     /note="synonym: SCD72A.10c"
CDS                  complement(6077. .7447)
                     /gene="SCO4124"
                     /note="SCD72A.10c, possible two-component sensor kinase,
                     len: 456aa; similar to many eg. TR:O86631 (EMBL:AL031155)
                     putative two-component sensor kinase from Streptomyces
                     coelicolor (429 aa) fasta scores; opt: 929, z-score:
                     995.1, E(): 0, 42.7% identity in 422 aa overlap. Contains
                     Pfam match to entry PF00512 signal, Histidine kinase and
                     possible membrane-spanning hydrophobic regions. Contains
                     possible hydrophobic membrane spanning regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative two-component sensor kinase"
                     /protein_id="CAB92377.1"
                     /db_xref="GI:8052436"
                     /db_xref="GOA:Q9KYE6"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="InterPro:IPR011712"
                     /db_xref="UniProt/TrEMBL:Q9KYE6"
                     /translation="MPATVIPGYYAPHLIYGPRTDARPARTPDHGRMEPVKPVPEDRA
                     AEPVTEYVQRVNRRIRAFDHRHPLLWDLTLTGFWVLAALVDGTGDGWRSVAHDPGVPV
                     GLVLVLSLALSVPVVWRRSHPRGALLLTLPAALVSAWSGAALQAALLQLLLVHHIALR
                     LPLRNLWWAVAVVAAPTCVSVARYGDSAGSWNQQLGSQLVAIGVAALIGITVRTRRGY
                     TEALEDRARRLETERDQQVRLAAAAERARIAREMHDIIGHNLSVITGLADGGRYAAAK
                     SPDRAAQALDAIGTTSRQALGELRRLLDVLRDDDQDGTGPEPELAPQPALTDLDRLLD
                     GVRRAGLPVRTTVKGEPALPEGRQLTLYRVIQEALTNTLKHAGPDATAHIELSYGSGG
                     AATLTVTDTGRTAPLPSRTPSGGRGLPGMRERTSLYGGTLEAGPRPAPEQGWRVHLHL
                     PEESPQ"
misc_feature         complement(6095. .6670)
                     /gene="SCO4124"
                     /note="Pfam match to entry PF00512 signal, Histidine
                     kinase, score -41.40, E-value 0.0087"
gene                 7473. .8006
                     /gene="SCO4125"
                     /note="synonym: SCD72A.11"
CDS                  7473. .8006
                     /gene="SCO4125"
                     /note="SCD72A.11, possible acetyltransferase, len: 177 aa;
                     weakly similar to many eg. TR:AAD52985 (EMBL:AF140221)
                     aminoglycoside 6'-N-acetyltransferase from Xanthomonas
```

```
                     maltophilia (153 aa) fasta scores; opt: 128, z-score:
                     164.9, E(): 0.1, 28.7% identity in 115 aa overlap.
                     Contains Pfam match to entry PF00583 Acetyltransf,
                     Acetyltransferase (GNAT) family."
                     /codon_start=1
                     /transl_table=11
                     /product="putative acetyltransferase"
                     /protein_id="CAB92378.1"
                     /db_xref="GI:8052437"
                     /db_xref="GOA:Q9KYE5"
                     /db_xref="InterPro:IPR000182"
                     /db_xref="UniProt/TrEMBL:Q9KYE5"
                     /translation="MIRSIHADEWPAVKELRLRALRDPVADLAYLETYEEAAARPDSY
                     WQERAARGAEGAAGAQQVVAVGPGGLWLGTLTVLVEEAGTTDWAGFPVERRQGHVVGV
                     FVRPEWRGSGLTKALFDAGLEWGWANGVERMRLIVHPGNTRAQGAYRKAGFAPSGRTV
                     PLEGGPGDHELEYVLER"
     misc_feature    7476. .7934
                     /gene="SCO4125"
                     /note="Pfam match to entry PF00583 Acetyltransf,
                     Acetyltransferase (GNAT) family, score 34.50, E-value
                     2.4e-06"
     gene            complement(8054. .9541)
                     /gene="SCO4126"
                     /note="synonym: SCD72A.12c"
     CDS             complement(8054. .9541)
                     /gene="SCO4126"
                     /note="SCD72A.12c, possible membrane protein, len: 495aa;
                     similar in regions to others from actinomycetes eg.
                     TR:Q9X8A2 (EMBL:AL049645) hypothetical protein from
                     Streptomyces coelicolor (427 aa) fasta scores; opt: 266,
                     z-score: 260.7, E(): 4.7e-07, 27.7% identity in 296 aa
                     overlap. Contains possible membrane-spanning hydrophobic
                     region."
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="CAB92379.1"
                     /db_xref="GI:8052438"
                     /db_xref="UniProt/TrEMBL:Q9KYE4"
                     /translation="MRPDDRRDQRPAQGGIPDGLLIGILAFLLGMTLLVWTATGLSAL
                     FAHGAWPDGVSFTGTPQAMRQLIAEPHDIPGAWPDAPEGALSGYGLFWGLFIGQLMVL
                     VVLTVFVMGTVARWRAVRQRHREQPTPAPAPAPPAAPVAQETPVQEPPPRHEVPTPRAP
                     HQDPALETEPPASAANGKRLGGWETHRAEGAVHYGPGESRHSTAVQTLRDAEGAALVI
                     TSNPALWSETKDARAKLGPTHLYDPNHLCDTPARVHWSPTGGCEDRQTATARAAALLT
                     PVRPTARLDQAIGDTAELLLRSYLHAAAIDGRTIRHVHRWSQGLQVQDAVRILRTHPK
                     AAPGSAGELEGSLTAHPERRDMAQQLTARALAALSTVNIREACTPNRTDALALDSFVH
                     EQGTLYVVGESIEDPRTSPGAMPLLTALVSSVVERGRRMAERSSSGRLDPPMTLVLDD
                     VAAVAPLPQLPDLLAGGTDAGMPTLALLRSREQGRARWPHDELPV"
     gene            complement(9615. .11030)
                     /gene="SCO4127"
                     /note="synonym: SCD72A.13c"
     CDS             complement(9615. .11030)

  Query Match             10.1%;  Score 91.4;  DB 1;  Length 299050;
  Best Local Similarity   46.8%;  Pred. No. 16;
  Matches  324;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172
              ||   | | |  |||||||| |  | | | || |   | |     ||||  |   | |||
Db      31174 CCTCACCCCCCTCGCGGTCGACCCCGCTCACCCCTTCCCGTACATCTCGGGCCTCTCCCT 31115

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232
              |  || | ||  | ||         |  ||    |  || | ||  |  |      | |
Db      31114 CAACCTCGCGGTCGTCGTCCGCAACCCCGTCACCGGCCACCGCCACTTCGCCCGCGTCAA 31055

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292
              | | | ||   |  ||  | ||  |  |||| |||  ||| ||| ||  |  | |||||
Db      31054 GGTCCCCCCACTGCTCTCCCGCTTCCTGGAGGCCTCCCCGGGCCGCTACGTCCCCGTCGA 30995

Qy        293 CGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGT 352
              ||||||     ||  | | |  ||   | |  | | | |  ||||  ||   |||
Db      30994 GGACGTCATCGCCGCCCACCTGGAGGAGCTGTTCCCGGGCATGGAGGTCCTGGAGCACCA 30935

Qy        353 GTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGC 412
              || || ||||  || |  || || |    |  | ||| |  || | || ||    |
Db      30934 CACCTTCCGCCTCACCCGCAACGAGGACCTGGAGGTCGAGGAGGACGACGCGGAGAACCT 30875

Qy        413 GCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGAT 472
              ||  | |||  | |   |  |  | |   |||||| ||| | | | ||   | |
Db      30874 GCTCCAGGCCCTGGAGAAGGAGCTGATGCGGCGCCGCCTGGGCCCGCCCGTCCGACTGGA 30815

Qy        473 CCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGA 532
                |  |||     || ||| ||   |||| || ||| ||     || || |  |   |
Db      30814 GGTCGAGGAGTCCGTCGACCGCGAGGTCCTGGACCTCCTCGTACGCGAGCTGAAGATCGG 30755

Qy        533 GGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGAC 592
              | ||| |  |  | | | | | ||   |   |  ||  |||| || | | || | ||
Db      30754 CGAGGCGGAGGTCTACCCGCTCCCCGGCCCCCTGGACCTCACCGGCCTCTTCCGCATCCA 30695

Qy        593 CCACATGATCAG---CAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCG 649
              |  |||   | |   | || |     || ||  |||||| |  |  ||| ||  ||
Db      30694 CAGCATCGACCGGCCCGAGCTGAAGTACCCGAAGTTCGTCGCCGGCACCCACCGCGACCT 30635

Qy        650 CCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCT 709
              || | ||       || |   ||  | ||||   |   | ||  |   || |
Db      30634 GGCCGAGGTCGAGTCCGCGTCCGCGCCGGACATCTTCGCCGCCCTGCGCACCAAGGACGT 30575

Qy        710 CGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGT 769
              | | || | | |  |  |  |  |  | | | | | | | | ||| ||  ||
Db      30574 CCTGCTCCACCACCCGTACGACTCCTTCTCCACCTCCGTGCAGGCGTTCCTGGAGCAGGC 30515

Qy        770 CGACGACCCCGCCGAGCTCCGCCTCGTCCTGCA 802
              || || |  | | ||  ||| |    ||  |||
Db      30514 CGCCGCCGACCCGGACGTCCTCGCGATCAAGCA 30482


RESULT 15
SCO939111/c
LOCUS       SCO939111             321250 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 8/29.
ACCESSION   AL939111 AL049497 AL096872 AL096884 AL109661 AL109663 AL109732
            AL109849 AL133219 AL133220 AL136500 AL355740 AL355752 AL445343
            AL590743 AL592262 AL645882
VERSION     AL939111.1  GI:24419015
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
   PUBMED   12000953
REFERENCE   2  (bases 1 to 321250)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 29, 2002 this sequence version replaced
            gi:20520791, gi:20520768, gi:20520769, gi:20520793, gi:20520794,
            gi:20520822, gi:20520770, gi:20520690, gi:20520671, gi:20520801,
            gi:20520696, gi:20520807, gi:20520866, gi:20520790.
FEATURES             Location/Qualifiers
     source          1. .321250
                     /organism="Streptomyces coelicolor A3(2)"
```

/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
gene complement(200. .2215)
/gene="SCO1964"
/note="synonym: SCC54.24c"
CDS complement(200. .2215)
/gene="SCO1964"
/note="SCC54.24c, possible export associated protein, len: 671aa; prediction based on the predicted function of surrounding genes and the short region (67-162aa) of similarity to TR:Q52357 (EMBL:L38824) terD, associated with phage inhibition, colicin resistance and tellurite resistance from Serratia marcescens IncHI2 plasmid R478 blast scores; E=2.1e-12, Identities = 40/98 (40%), Positives = 55/98 (56%) which may represent an active site."
/codon_start=1
/transl_table=11
/product="putative export associated protein"
/protein_id="CAB38154.1"
/db_xref="GI:4468700"
/db_xref="GOA:Q9Z502"
/db_xref="InterPro:IPR003325"
/db_xref="UniProt/TrEMBL:Q9Z502"
/translation="MTAELTRGQNHPLPRARLEIRVSAGTPVVAGATLGDENGTIHGV
ERVVHPGAPTLPGLEVSRQAAADHRLAVDLDAVPDAVHRVSVLLALPAGGQGPARFGA
VAAPFVAVTDLDGDEVASYTITGLEAESAVVALELYRRQGAWKVRAVGQGYAGGLAEL
LADQKLSQAHQLAATIHEAVASGLARSIPAPPAAAPAHRPDHGAMPGAGPGGPVPPAS
PYDPQGPSAPGPQTPGAPGRPGGQQPYPSGPGEPATAGQPSAPAAGGPIDYSHPRRQS
AAPPPPPPAAPASEPGRPARPVAGDATGWSMEERLYNQVWGMFEDLARTTAAYRSAVD
FADSRMEKELDQVLSDPRSRIGGQGDAAREAARARHGQLVSQAREVLDRDVAQLVAEA
EVVEPALPAAFARWDNPVWHAYRVPMEIPMALRLGDLHLPEADRIRIPMLIRLPLERG
LWIDSGRSASLDGSFADSHEMRRLGLETAVSHAARLLAVYPAGEFTVHVIDPAGSGAQ
ALAPLAQSGVLAAPPAQGAAGAADVLARLTQRVDLVQMALRGGAPDALPPGLDTSQQL
LIVNDFPHGFDDRAVNQLRYLADEGPAVGVHLMMVADREESAGFGPLLDPLWRSLLRL
TPVADDHLADPWVGHAWTYEPSLVPPGSQVLQQVLAQVAAARRSWDR"
gene complement(2340. .2918)
/gene="SCO1965"
/note="synonym: SCC54.25c"
CDS complement(2340. .2918)
/gene="SCO1965"
/note="SCC54.25c, possible export associated protein, len: 192aa; similar to many associated with plasmid borne resistance export systems egs. TR:Q52353 (EMBL:L38824) from the IncHI2 plasmid R478 of Serratia marcescens (193 aa) fasta scores; opt: 497, z-score: 593.2, E(): 9.8e-26, (42.2% identity in 192 aa overlap) and SW:TERD_ALCSP terD, tellurium resistance protein from the pMER610 plasmid of Alcaligenes sp. (192 aa) fasta scores; opt: 357, z-score: 429.6, E(): 1.3e-16, (37.2% identity in 188 aa overlap)."
/codon_start=1
/transl_table=11
/product="putative export associated protein"
/protein_id="CAB38155.1"
/db_xref="GI:4468701"
/db_xref="GOA:Q9Z501"
/db_xref="InterPro:IPR003325"
/db_xref="UniProt/TrEMBL:Q9Z501"
/translation="MTVNMTKGQAISLQKSDGGSLTSVRMGLGWQAAPRRGLFGSRTR
EIDLDASAVLFADKQPVDVVFFRHLVSDDGSVRHTGDNLVGGVGQGGDDEAILVDLQR
VPVHIDQIVFTVNSFTGQTFQEVQNAFCRLVDETNGQELARYTLAGGGAFTAQIMAKV
HRAGQGWSMTAIGTPANGRTFQDLMPAILPVL"
gene complement(3061. .5199)
/gene="SCO1966"
/note="synonyms: SC3C9.01c, SCC54.26c, uvrB"
CDS complement(3061. .5199)
/gene="SCO1966"
/note="SCC54.26c, partial CDS, probable uvrB, ABC excision nuclease subunit B, len: >113aa; similar to many eg. SW:UVRB_MICLU uvrB, ABC excision nuclease subunit B from Micrococcus luteus (709 aa) fasta scores; opt: 248, z-score: 280.5, E(): 2.6e-08, (45.9% identity in 111 aa overlap).
SC3C9.01c, uvrB, excinuclease ABC subunit B (fragment), len: >635 aa; highly similar to SW:UVRB_BACSU (EMBL:AF017113) Bacillus subtilis excinuclease ABC subunit B (DinA protein) UvrB or DinA or Uvr aa; fasta scores: opt: 2591 Z-score: 2687.6 bits: 507.5 E(): 4.3e-142; 63.667% identity in 600 aa overlap. Contains Pfam matches to entries PF00270 DEAD, DEAD/DEAH box helicase and PF00271 helicase_C, Helicase conserved C-terminal domain and match to Prosite entry PS00017 ATP/GTP-binding site motif A (P-loop). Also contains a possible coiled-coil region at approx. residues 273. .290"
/codon_start=1
/transl_table=11
/product="ABC excision nuclease subunit B"
/protein_id="CAD55287.1"
/db_xref="GI:24419016"
/db_xref="GOA:Q8CK11"
/db_xref="InterPro:IPR001410"
/db_xref="InterPro:IPR001650"
/db_xref="InterPro:IPR001943"
/db_xref="InterPro:IPR004807"
/db_xref="InterPro:IPR006935"
/db_xref="InterPro:IPR011545"
/db_xref="UniProt/TrEMBL:Q8CK11"
/translation="MRPVSQIERTVAPFEVVSPYQPSGDQPTAIAELARRVQAGEKDV
VLLGATGTGKSATTAWMIEKLQRPTLVMAPNKTLAAQLANEFRELLPNNAVEYFVSYY
DYYQPEAYVPQSDTYIEKDSSINEEVERLRHSATNSLLTRRDVIVVASVSCIYGLGTP
QEYVDRMVPLRVGEEHDRDELLRRFVDIQYTRNDMAFARGTFRVRGDTIEIFPVYEEL
AVRIEMFGDEIEALSTLHPVTGEIISEDQQLYVFPASHYVAGPERLERAVNDIEKELA
ERLTELEKQGKLLEAQRLRMRTTYDIEMLRQIGSCSGVENYSMHFDGRSPGSPPNTLL
DYFPDDFLLVIDESHVTVPQIGAMYEGDASRKRTLVDHGFRLPSALDNRPLKWEEFQE
RIGQTVYLSATPGAYELSRSDGAVEQIIRPTGLVDPEVVVKPTEGQIDDLVHEIRRRT
EKDERVLVTTLTKKMAEDLTDYFVELGIQVRYLHSDVDTLRRVELLRELRAGEYDVLV
GINLLREGLDLPEVSLVAILDADKEGFLRSGTSLIQTIGRAARNVSGQVHMYADKITP
AMEKAIDETNRRREKQVAFNKANGVDPQPLRKKINDIVAQIAREDVDTEQLLGSGYRQ
TKEGKGAKAPVPALGGQKTGGAKAARGRAKETAVTDRPAAELAEQIEDLTTRMRAAAA
DLQFEIAARLRDEVSEMKKELRQMREAGLA"
misc_feature complement(3541. .3801)
/gene="SCO1966"
/note="Pfam match to entry PF00271 helicase_C, Helicase conserved C-terminal domain, score 75.00, E-value 1.5e-18"
misc_feature complement(4753. .5166)
/gene="SCO1966"
/note="Pfam match to entry PF00270 DEAD, DEAD/DEAH box helicase, score 27.00, E-value 1.2e-07"
misc_feature complement(5035. .5058)
/gene="SCO1966"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
gene complement(5239. .6117)
/gene="SCO1967"
/note="synonym: SC3C9.02c"
CDS complement(5239. .6117)
/gene="SCO1967"
/note="SC3C9.02c, possible integral membrane protein, len: 292 aa; similar to TR:Q9KZI4 (EMBL:AL353863) Streptomyces coelicolor putative integral membrane protein SCG8A.15c, 244 aa; fasta scores: opt: 703 Z-score: 726.3 bits: 142.1 E(): 7.6e-33; 48.430% identity in 223 aa overlap. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAC42741.1"
/db_xref="GI:14571767"
/db_xref="InterPro:IPR005330"
/db_xref="UniProt/TrEMBL:Q93JJ4"
/translation="MHSTVQGFGYGLVTPLVAYLMACLGGALGLRCTARAVLVSRSWR
PGWLALGSAAIGSGIWTMHFVAMMGFTIEHTPIRYDWLMTFASLAVAIVMVGVGIFIV
GYRGARGTALFTGGTITGLGIASMHYLGMAGMHLDGQLTYNTFTVAVSVVIAMVAATA
ALWAAGQVRGFLWSVGAALIMGLAVTGMHYTGMAAVEVHLSGTAEPSVGGSPAELLAP
MLIGPLAFLLLAGVVVMFDPLMVMGRPARVPAERKPGIPAHTEVPRTIRHPAHHRPRR
TRHPLGHRRSRTPQNR"

```
     RBS             6399. .6403
     gene            6409. .7278
                     /gene="SCO1968"
                     /note="synonym: SC3C9.03"
     CDS             6409. .7278
                     /gene="SCO1968"
                     /note="SC3C9.03, probable secreted hydrolase, len: 289 aa;
                     similar to SW:GLPQ_BACSU (EMBL:Z26522) Bacillus subtilis
                     glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)
                     GlpQ, 293 aa; fasta scores: opt: 495 Z-score: 576.9 bits:
                     114.7 E(): 1.6e-24; 34.114% identity in 299 aa overlap.
                     Contains possible cleavable N-terminal region signal
                     peptide sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secreted hydrolase"

  Query Match             10.1%;  Score 91;  DB 1;  Length 321250;
  Best Local Similarity   45.8%;  Pred. No. 17;
  Matches  353;  Conservative    0;  Mismatches  415;  Indels    3;  Gaps    1;

Qy         81 CCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCG 140
Db     186917 CCGGGGGTGGCGCGGGCGGGCCGCAGCGGCACCCCGCACCGCCGGGCTGCGAACCCTCCT 186858

Qy        141 CGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATC 200
Db     186857 CCCCACCGGCACCCCGGCACCGCCGACGCACCGGTACCGTTCACCCCCATGCGCACCCTC 186798

Qy        201 GAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGG 260
Db     186797 CTGATCACGGGCCCCGGCGGCACCGGCCGTACCACCGTCGCCGCCGCCACCGCGCTCACC 186738

Qy        261 GAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGC 320
Db     186737 GCGGCCCGGCAGGGCACCCGCACCCTCGTCCTCGGAACCGACCGCGACGACACCCTCGGC 186678

Qy        321 CCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCC 380
Db     186677 GCCGCACTCGGCGTGCGCACCGGCCCGGCCCCGACCTCCGTCGAACCCGGCCTCACCGCT 186618

Qy        381 GACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACC 440
Db     186617 TGGCGCCCCGACGCCGCCCAGGGCTTCCGGGACGCCCTCGCCGCCCTCCAGGACCGCGCC 186558

Qy        441 CTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTC 500
Db     186557 GCCTCCGCGCTCGACCTCCTCGGCGCCGCCCGCCTCGACCCCCAGGAACTCACCCCGCTG 186498

Qy        501 CTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTG 560
Db     186497 CCCGGGGCCGACGACCTCGCCCTGCTGCGTGCTCTGCGCGAGGCCGCCCTCGCCGAGGCG 186438

Qy        561 GCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACG 620
Db     186437 CACGACCTCCTCGTCGTCGACCTGCCGCCCGCCCCGCGCGCCCTCGCCCTGCTCGCCGCC 186378

Qy        621 CGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGC 680
Db     186377 CCCGAGGAACTGCGCCGCGCCCTGCGCCGCCTGCTCCCGCCCGAGCGGCAGGCCGCCCGC 186318

Qy        681 CTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTC 740
Db     186317 GCCCTGCGCCCGGTCCTCGGCCG---GCTCGCCGGCGTCCCCATGCCGACCGAGGCGCTG 186261

Qy        741 GACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTG 800
Db     186260 TACGAGGCCGCCGCCCGCTGGGACCTGGAACTCGCCGCCGCCGAATCCGTCCTGGCCGAC 186201

Qy        801 CACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGGGAGGTC 851
Db     186200 CGGAACACCGTCGTACGGCTGGTCGCCGAGCCCGGACCCGCCGGAGCCGAC 186150
```

```
Search completed: April 28, 2006, 02:39:41
Job time : 4995.51 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 434.473 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_75000_75900
Perfect score: 901
Sequence: 1 acggccgggacgcggacgac.....gaggaggatcgcgccgcgga 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	899.4	99.8	76994	12	ADP64454	Adp64454 Sorangium
	2	91.4	10.1	3780	6	ABN85142	Abn85142 Streptomy
	3	90.2	10.0	1433	13	ADX33042	Adx33042 Plant ful
	4	90.2	10.0	1456	13	ADX32774	Adx32774 Plant ful
c	5	88.4	9.8	77536	3	AAA14651	Aaa14651 Nucleotid
	6	87.8	9.7	22533	13	ADQ91713	Adq91713 Polyketid
	7	87.8	9.7	22533	13	AEB87005	Aeb87005 Streptomy
	8	87.8	9.7	22533	13	AEB86802	Aeb86802 Streptomy
	9	87.8	9.7	164051	13	ADQ91695	Adq91695 Polyketid
	10	87.8	9.7	164051	13	AEB86862	Aeb86862 Streptomy
	11	87.8	9.7	164051	13	AEB86659	Aeb86659 Streptomy
	12	87.4	9.7	3948	8	ACA37866	Aca37866 Prokaryot
	13	87	9.7	1254	13	ADX32903	Adx32903 Plant ful
	14	87	9.7	1434	13	ADX32964	Adx32964 Plant ful
	15	87	9.7	9222	8	AAL61171	Aal61171 Actinosyn
c	16	87	9.7	82746	8	AAL61224	Aal61224 Actinosyn
	17	86.8	9.6	27541	4	AAD17185	Aad17185 Streptomy
	18	86.8	9.6	125401	4	AAD17186	Aad17186 Streptomy
	19	86.2	9.6	31263	8	ACA37577	Aca37577 Prokaryot
	20	85.4	9.5	58857	3	AAA58471	Aaa58471 Nucleotid
	21	84	9.3	1047	10	ABZ66716	Abz66716 Orthosomy
	22	84	9.3	45055	10	ABZ66808	Abz66808 Orthosomy
	23	83.8	9.3	65140	4	AAD17184	Aad17184 Streptomy
	24	83.4	9.3	1839	13	ADX30727	Adx30727 Plant ful
c	25	82.8	9.2	138203	14	ADZ51725	Adz51725 FR-008 po
	26	82.2	9.1	1380	10	ADG33879	Adg33879 Actinomyc
	27	82.2	9.1	113193	8	AAD54645	Aad54645 Streptomy
	28	82	9.1	4770	8	AAL61172	Aal61172 Actinosyn
	29	82	9.1	104096	13	ADX56092	Adx56092 Streptomy
	30	81.8	9.1	5802	6	ABS78696	Abs78696 S. kaniha
	31	81.8	9.1	5802	14	AEB12688	Aeb12688 Enediyne
	32	81.8	9.1	5802	14	AEC10281	Aec10281 Streptomy
	33	81.6	9.1	2312	2	AAQ35141	Aaq35141 srmR codi
	34	81.6	9.1	7788	10	ADI23894	Adi23894 Streptomy
	35	81.6	9.1	37360	10	ADI23892	Adi23892 Streptomy
	36	81	9.0	936	10	ADC16609	Adc16609 Secondary
	37	81	9.0	936	13	ADR13666	Adr13666 Amidase c
	38	81	9.0	14055	8	AAL61170	Aal61170 Actinosyn
	39	80.6	8.9	1266	8	ACA27164	Aca27164 Prokaryot
	40	80	8.9	1475	13	ADX47828	Adx47828 Plant ful
	41	80	8.9	1478	13	ADX50159	Adx50159 Plant ful
	42	80	8.9	1968	12	ADM80080	Adm80080 Spiramyci
	43	80	8.9	1968	12	ADN97596	Adn97596 S ambofac
	44	80	8.9	30943	12	ADM80034	Adm80034 Spiramyci
	45	80	8.9	30943	12	ADN97550	Adn97550 S ambofac

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

PS Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC The present invention describes an isolated recombinant polynucleotide
CC (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC synthase (PKS) protein or a fragment comprising one or more domains of
CC the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC host cell (III) comprising (II), or (I) integrated into the cell
CC chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC the native disorazole PKS by the inactivation of one or more domain of
CC the disorazole PKS and/or addition of one or more domain of a non-
CC disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC expression system (VI) capable of producing a disorazole synthase domain
CC in a host cell, where the system comprises an encoding sequence for a
CC disorazole polyketide synthase domain, and the encoding sequence is
CC operably linked to control sequences effective in the cell to produce RNA
CC that is translated into the domain; (7) host cell modified to contain
CC (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC eliminate production of disorazole; (9) an isolated polypeptide encoded
CC by (I); (10) hybrid polyketide synthase comprising one or more
CC polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC of at least 200 base pairs with a sequence identical or substantially
CC identical to a protein encoding region of the 77294 base pair sequence of
CC SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC (I) is useful for producing disorazole polyketides. (III) is useful for
CC producing a polyketide, which involves growing (III) under conditions,
CC where a polyketide synthesised by a PKS comprising a protein encoded by
CC the polynucleotide molecule is produced in the cell. (I) is useful for
CC producing libraries of PKSs, where the polyketide producing colonies are
CC identified and isolated, and the produced polyketides are used
CC collectively in a panel to represent a library or assessed individually
CC for activity. The disorazole polyketide produced by (I) is useful in
CC molecular biology, chemistry, recombinant DNA technology, human and
CC veterinary medicine and agricultural applications. The disorazole
CC polyketide produced by (I) is formulated as a pharmaceutical composition,
CC useful for its activity against microorganisms, in a mammal. The present
CC sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

Query Match 99.8%; Score 899.4; DB 12; Length 76994;
Best Local Similarity 99.9%; Pred. No. 7.9e-124;
Matches 900; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ACGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74700 ACGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGC 74759

Qy         61 TCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74760 TCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGC 74819

Qy        121 ACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74820 ACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACC 74879

Qy        181 CGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74880 CGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCC 74939

Qy        241 ACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74940 ACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCC 74999

Qy        301 AGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75000 AGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCC 75059

Qy        361 TCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75060 TCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGG 75119

Qy        421 CGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75120 CGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGG 75179

Qy        481 ACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75180 ACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGC 75239

Qy        541 CGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75240 CGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGA 75299

Qy        601 TCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCC 660
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      75300 TCAGCAAGATCGGCGACGCGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCC 75359

Qy        661 TCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75360 TCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATC 75419

Qy        721 CCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75420 CCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCG 75479

Qy        781 CCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75480 CCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCG 75539

Qy        841 GCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75540 GCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCGG 75599

Qy        901 A 901
              |
Db      75600 A 75600
```

RESULT 2
ABN85142
ID ABN85142 standard; DNA; 3780 BP.
XX
AC ABN85142;
XX
DT 16-SEP-2002 (first entry)
XX
DE Streptomyces coelicolor polyphosphate kinase, ppk, gene.
XX
KW Antibiotic; anticancer; anthelmintic; antifungal; herbicidal;
KW insecticidal; polyphosphate kinase; enzyme; ppk; gene; ds.
XX
OS Streptomyces coelicolor.
XX
PN FR2818991-A1.
XX
PD 05-JUL-2002.
XX
PF 04-JAN-2001; 2001FR-00000088.
XX
PR 04-JAN-2001; 2001FR-00000088.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Virolle MN, Chouayekh H;
XX
DR WPI; 2002-522185/56.
XX
PT Increasing production of microbial metabolites, useful e.g. for
PT antibiotic preparation, using strains in which the polyphosphate kinase
PT gene is inactivated.

```
XX
PS   Claim 15; Fig 1; 32pp; French.
XX
CC   The present invention relates to methods using microbial strains in which
CC   the ppk (polyphosphate kinase) gene is inactivated for increasing
CC   production of compounds of interest. The present sequence is the ppk gene
CC   from Streptomyces coelicolor, which was used to illustrate the invention.
CC   The microbial strains are useful for the production of metabolites with
CC   antibiotic (particularly), anticancer, anthelmintic, antifungal,
CC   herbicidal, insecticidal or enzymatic activity. Strains that cannot
CC   produce a functional PPK enzyme overproduce metabolites. PPK is essential
CC   for accumulation of phosphate reserves, so cells that lacking PPK soon
CC   become deficient in inorganic phosphate and this induces overexpression
CC   of metabolites, at up to 20-30 times the normal level
XX
SQ   Sequence 3780 BP; 562 A; 1496 C; 1274 G; 448 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 91.4;  DB 6;  Length 3780;
  Best Local Similarity   46.8%;  Pred. No. 4.5e-05;
  Matches  324;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172
              ||   | | |  |||||||| |  | | | || |   | |     ||||  |   | |||
Db       1666 CCTCACCCCCCTCGCGGTCGACCCCGCTCACCCCTTCCCGTACATCTCGGGCCTCTCCCT 1725

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232
              |  || | ||  | ||         |  ||   |   || | ||  | |     |  |
Db       1726 CAACCTCGCGGTCGTCGTCCGCAACCCCGTCACCGGCCACCGCCACTTCGCCCGCGTCAA 1785

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292
              | | | ||   |  ||  | ||  |  |||| |||  ||| ||| ||  |  | |||||
Db       1786 GGTCCCCCCACTGCTCTCCCGCTTCCTGGAGGCCTCCCCGGGCCGCTACGTCCCCGTCGA 1845

Qy        293 CGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGT 352
              ||||||     ||  | | |  ||   | |  | | | |  ||||  ||   |||
Db       1846 GGACGTCATCGCCGCCCACCTGGAGGAGCTGTTCCCGGGCATGGAGGTCCTGGAGCACCA 1905

Qy        353 GTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGC 412
              || || ||||  || |  || || |   |  | ||| |  || | || ||    |
Db       1906 CACCTTCCGCCTCACCCGCAACGAGGACCTGGAGGTCGAGGAGGACGACGCGGAGAACCT 1965

Qy        413 GCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGAT 472
              ||  | |||  | |   |  |  |  |    |||||| ||| | | | ||   | |
Db       1966 GCTCCAGGCCCTGGAGAAGGAGCTGATGCGGCGCCGCCTGGGCCCGCCCGTCCGACTGGA 2025

Qy        473 CCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGA 532
              |  |||      || ||| ||   |||| || ||| ||      || ||| |   |   |
Db       2026 GGTCGAGGAGTCCGTCGACCGCGAGGTCCTGGACCTCCTCGTACGCGAGCTGAAGATCGG 2085

Qy        533 GGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGAC 592
              | || |  |  | | | | ||   |  |  ||  ||||| || | || | ||
Db       2086 CGAGGCGGAGGTCTACCCGCTCCCCGGCCCCCTGGACCTCACCGGCCTCTTCCGCATCCA 2145

Qy        593 CCACATGATCAG---CAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCG 649
              |  |||   | |   | || |     || ||   ||||||| |  |  ||| ||   ||
Db       2146 CAGCATCGACCGGCCCGAGCTGAAGTACCCGAAGTTCGTCGCCGGCACCCACCGCGACCT 2205

Qy        650 CCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCT 709
              || | ||        || |    ||   | ||||    |   | ||  |    || |
Db       2206 GGCCGAGGTCGAGTCCGCGTCCGCGCCGGACATCTTCGCCGCCCTGCGCACCAAGGACGT 2265

Qy        710 CGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGT 769
              |  | || | | |        |   |  | | |  |  | | |||  ||   ||
Db       2266 CCTGCTCCACCACCCGTACGACTCCTTCTCCACCTCCGTGCAGGCGTTCCTGGAGCAGGC 2325

Qy        770 CGACGACCCCGCCGAGCTCCGCCTCGTCCTGCA 802
              || || |  | | ||  ||| |    ||  |||
Db       2326 CGCCGCCGACCCGGACGTCCTCGCGATCAAGCA 2358


RESULT 3
ADX33042
ID   ADX33042 standard; cDNA; 1433 BP.
XX
AC   ADX33042;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polynucleotide seqid 15862.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content; gene; ss.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 15862; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC   of the invention are also useful in physical arrays of molecules and as
CC   plant breeding markers. The recombinant DNA construct is useful for
CC   improving plant tolerance to cold, heat, drought, herbicides, extreme
CC   osmotic conditions, pathogens or pests, for manipulating growth rate in
CC   plant cells by modification of the cell cycle pathway, for conferring
CC   increased resistance to plant disease, for producing galactomannan,
CC   lignin or plant growth regulators, for increasing the rate of homologous
CC   recombination in plants, for improving yield by modification of
CC   photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC   or by providing improved plant growth and development under at least one
CC   stress condition or for modifying seed oil or protein yield and/or
CC   content. This sequence represents a plant full length insert
CC   polynucleotide that can be used in the recombinant DNA construct of the
CC   invention.
XX
SQ   Sequence 1433 BP; 254 A; 492 C; 452 G; 235 T; 0 U; 0 Other;

  Query Match             10.0%;  Score 90.2;  DB 13;  Length 1433;
  Best Local Similarity   46.8%;  Pred. No. 7.8e-05;
  Matches  407;  Conservative    0;  Mismatches  438;  Indels   24;  Gaps    3;

Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
              ||   | | |  |||||  | | || |   ||  |||||              ||  | |
Db        275 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 334
```

```
Qy          66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
               |   |  | |||     || |  | || |  || |  || || |  | | | ||  | |
Db         335 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 394

Qy         126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
               ||  ||| |  | |||   ||   ||||||| ||  ||||| | |  |  | |||    |
Db         395 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 451

Qy         186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
               || | |   |  |||||   ||   |   |||  |||  |   || || | |
Db         452 CTGCTGAGCCAGCTCGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 511

Qy         246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
                | ||| |  ||  |   || | |||| ||          | ||||  ||  |  ||||
Db         512 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 571

Qy         306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
               | |||  |     | |   || | || | || ||| ||| ||  |              |
Db         572 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC-----------C 619

Qy         366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
               || ||| ||||  |  | |||    || ||| | || |  |||||||||   |||| ||
Db         620 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 679

Qy         426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
                    |||| || |  |  | || |  ||  || | |  |  || ||  |||| ||||
Db         680 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 730

Qy         486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
               ||  ||| ||    |||| |  | |  | |||| |  |  || |   |  |     |  |
Db         731 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 790

Qy         546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
                  ||    |  |||||    |  |      | ||||||    | |||       ||
Db         791 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 850

Qy         606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
                  ||    || |     | | |  | |        || | | || || |    | |
Db         851 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 910

Qy         666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
               |   | |||  ||       | |||  |||  |  ||     | ||||    | | |||  |
Db         911 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCCCGCAAGCCCCTC 970

Qy         726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
               |     ||      |||| | |  || | |  |  |||||  ||   ||  ||||| |
Db         971 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCGCACGCGCACGGCGCCGTG 1030

Qy         786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845
               ||| |||    |  | |||  ||| | ||  |    | | |||| |   ||| |    |
Db        1031 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1090

Qy         846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874
               | || |     ||   | | |  | ||||
Db        1091 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1119
```

RESULT 4
ADX32774
ID ADX32774 standard; cDNA; 1456 BP.
XX
AC ADX32774;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polynucleotide seqid 15594.
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX
DR WPI; 2004-180133/17.
XX
PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 15594; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 1456 BP; 262 A; 498 C; 455 G; 241 T; 0 U; 0 Other;

Query Match 10.0%; Score 90.2; DB 13; Length 1456;
Best Local Similarity 46.8%; Pred. No. 7.8e-05;
Matches 407; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```
Qy           6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
               ||   | | |  |||||  | | || |  ||  |||||           ||  |  |
Db         284 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 343

Qy          66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
               |   | | |||      || |  | || |  || |  || || |  | | | ||  | |
Db         344 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 403

Qy         126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
               ||  ||| |  | |||   ||   ||||||| ||  ||||| | |  |  | |||    |
Db         404 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 460

Qy         186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
               || | |   |  ||||||    ||     |  |||  |||  |    || || | |
Db         461 CTGCTGAGCCAGCTCGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 520
```

```
Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
Db        521 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 580

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
Db        581 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC----------C 628

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
Db        629 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 688

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
Db        689 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 739

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
Db        740 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 799

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
Db        800 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 859

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
Db        860 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 919

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
Db        920 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCCCGCAAGCCCCTC 979

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
Db        980 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCGCACGCGCACGGCGCCGTG 1039

Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845
Db       1040 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1099

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874
Db       1100 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1128
```

```
RESULT 5
AAA14651/c
ID   AAA14651 standard; DNA; 77536 BP.
XX
AC   AAA14651;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW   FK-506 binding protein; polyketide compound; transplant rejection;
KW   graft-versus-host disease; uveitis; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(412. .1836)
FT                   /*tag= a
FT                   /note= "fkbW gene"
FT   CDS             complement(2020. .3579)
FT                   /*tag= b
FT                   /note= "fkbV gene"
FT   CDS             3969. .4496
FT                   /*tag= c
FT                   /note= "fkbR2 gene"
FT   CDS             complement(4595. .5488)
FT                   /*tag= d
FT                   /note= "fkbR1 gene"
FT   CDS             5601. .6818
FT                   /*tag= e
FT                   /note= "fkbE gene"
FT   CDS             6808. .8052
FT                   /*tag= f
FT                   /note= "fkbF gene"
FT   CDS             8156. .8824
FT                   /*tag= g
FT                   /note= "fkbG gene"
FT   CDS             complement(9122. .9883)
FT                   /*tag= h
FT                   /note= "fkbH gene"
FT   CDS             complement(9894. .10994)
FT                   /*tag= i
FT                   /note= "fkbI gene"
FT   CDS             complement(10987. .11247)
FT                   /*tag= j
FT                   /note= "fkbJ gene"
FT   CDS             complement(11244. .12092)
FT                   /*tag= k
FT                   /note= "fkbK gene"
FT   CDS             complement(12113. .13150)
FT                   /*tag= l
FT                   /note= "fkbL gene"
FT   CDS             complement(13212. .23988)
FT                   /*tag= m
FT                   /note= "fkbC gene"
FT   misc_feature    complement(13452. .13662)
FT                   /*tag= n
FT                   /note= "ACP6"
FT   misc_feature    complement(13761. .14394)
FT                   /*tag= o
FT                   /note= "KR6"
FT   misc_feature    complement(14517. .15294)
FT                   /*tag= p
FT                   /note= "ER6"
FT   misc_feature    complement(15438. .16587)
FT                   /*tag= q
FT                   /note= "dehydratase domain (DH) 6"
FT   misc_feature    complement(16587. .17820)
FT                   /*tag= r
FT                   /note= "acyltransferase domain (AT) 6"
FT   misc_feature    complement(17820. .19053)
FT                   /*tag= s
FT                   /note= "KS6"
FT   misc_feature    complement(19116. .19326)
FT                   /*tag= t
FT                   /note= "ACP5"
FT   misc_feature    complement(19464. .20097)
FT                   /*tag= u
FT                   /note= "KR5"
FT   misc_feature    complement(20241. .21420)
FT                   /*tag= v
FT                   /note= "DH5"
FT   misc_feature    complement(21420. .22653)
FT                   /*tag= w
FT                   /note= "AT5"
FT   misc_feature    complement(22653. .23892)
FT                   /*tag= x
FT                   /note= "KS5"
FT   CDS             complement(23992. .46573)
FT                   /*tag= y
FT                   /note= "fkbB gene"
FT   misc_feature    complement(24163. .24373)
FT                   /*tag= z
```

```
FT                   /note= "ACP4"
FT   misc_feature    complement(24997. .26146)
FT                   /*tag=  aa
FT                   /note= "DH4 (inactive)"
FT   misc_feature    complement(26146. .27430)
FT                   /*tag=  ab
FT                   /note= "AT4"
FT   misc_feature    complement(27430. .28684)
FT                   /*tag=  ac
FT                   /note= "KS4"
FT   misc_feature    complement(28750. .28960)
FT                   /*tag=  ad
FT                   /note= "ACP3"
FT   misc_feature    complement(28750. .28960)
FT                   /*tag=  al
FT                   /note= "DH2 (inactive)"
FT   misc_feature    complement(29092. .29740)
FT                   /*tag=  ae
FT                   /note= "KR3"
FT   misc_feature    complement(29869. .31018)
FT                   /*tag=  af
FT                   /note= "DH3 (inactive)"
FT   misc_feature    complement(31018. .32185)
FT                   /*tag=  ag
FT                   /note= "AT3"
FT   misc_feature    complement(32185. .33439)
FT                   /*tag=  ah
FT                   /note= "KS3"
FT   misc_feature    complement(33505. .33715)
FT                   /*tag=  ai
FT                   /note= "ACP2"
FT   misc_feature    complement(33823. .34480)
FT                   /*tag=  aj
FT                   /note= "KR2"
FT   misc_feature    complement(34606. .35749)
FT                   /*tag=  ak
FT   misc_feature    complement(35749. .37144)
FT                   /*tag=  am
FT                   /note= "AT2"
FT   misc_feature    complement(37145. .38296)
FT                   /*tag=  an
FT                   /note= "KS2"
FT   misc_feature    complement(38371. .38581)
FT                   /*tag=  ao
FT                   /note= "ACP1"
FT   misc_feature    complement(38677. .39307)
FT                   /*tag=  ap
FT                   /note= "KR1"
FT   misc_feature    complement(39442. .40609)
FT                   /*tag=  aq
FT                   /note= "DH1"
FT   misc_feature    complement(40609. .41842)
FT                   /*tag=  ar
FT                   /note= "AT1"
FT   misc_feature    complement(41842. .43093)
FT                   /*tag=  as
FT                   /note= "KS of extender module 1 (KS1)"
FT   misc_feature    complement(43144. .43660)
FT                   /*tag=  at
FT                   /note= "ACP of loading domain"
FT   misc_feature    complement(43777. .44629)
FT                   /*tag=  au
FT                   /note= "ER of loading domain"
FT   misc_feature    complement(44974. .46573)
FT                   /*tag=  av
FT                   /note= "CoA ligase of laoding domain"
FT   CDS             46754. .47788
FT                   /*tag=  aw
FT                   /note= "fkbO gene"
FT   CDS             47785. .52272
FT                   /*tag=  ax
FT                   /note= "fkbP gene"
FT   CDS             52275. .71465
FT                   /*tag=  ay
FT                   /note= "fkbA gene"
FT   misc_feature    52362. .53576
FT                   /*tag=  az
FT                   /note= "KS7"
FT   misc_feature    53577. .54716
FT                   /*tag=  bb
FT                   /note= "AT7"
FT   misc_feature    54717. .55871
FT                   /*tag=  ba
FT                   /note= "DH7"
FT   misc_feature    56019. .56819
FT                   /*tag=  bc
FT                   /note= "ER7"
FT   misc_feature    56943. .57575
FT                   /*tag=  bd
FT                   /note= "KR7"
FT   misc_feature    57710. .57920
FT                   /*tag=  be
FT                   /note= "ACP7"
FT   misc_feature    57990. .59243
FT                   /*tag=  bf
FT                   /note= "KS8"
FT   misc_feature    59244. .60398
FT                   /*tag=  bg
FT                   /note= "AT8"
FT   misc_feature    60399. .61412
FT                   /*tag=  bh
FT                   /note= "DH8 (inactive)"

  Query Match             9.8%;  Score 88.4;  DB 3;  Length 77536;
  Best Local Similarity   45.5%;  Pred. No. 7.7e-05;
  Matches  314;  Conservative    0;  Mismatches  376;  Indels    0;  Gaps    0;

Qy        124 ACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGT 183
              |||   ||||||  | || || |   ||  ||  ||  |           ||| |  | ||
Db      44340 ACGGCCTCGTCGACGCGGGCGAGCTGCGACCCGGCCAGAAGGTCCTGGTCCACGCCGCGA 44281

Qy        184 CCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACG 243
              |   | | |    ||| | | ||  ||    || | |||   || ||||   |    |||
Db      44280 CGGGCGGTGTCGGCATGGCGGCGACCCGGATCGCCCGCCACCTCGGCGCCGAGGTGTACG 44221

Qy        244 TCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGC 303
              |  | | | | ||   | |     | ||  |||   ||   | ||| |  || |||
Db      44220 CCACCGCCAGTCCCGCCAAACAACACCTGCTGCACGCCGACGGCTTCGACACCGACCATG 44161

Qy        304 GCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCC 363
              | |  || |    || | |    || |||     ||     ||  |   ||   |||
Db      44160 TGGCGAACTCCCGCAGCGCCGCGTTCGCCGACACCTTCCCGCCGGTCGACGTCGTGCTCA 44101

Qy        364 TCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGA 423
              |   | |  ||  |    | ||| ||    |   |||  |  |||  || |  |   |
Db      44100 ACTCGCTCACCGGTGAACTCCTCGACGCGTCCATCGGCCTGCTCGCACCGGGCGGCCGGT 44041

Qy        424 TCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACG 483
              ||| || |     ||    || |  || |  ||||  ||   | | | ||  |  || ||
Db      44040 TCGTCGAGATGGGCAAGACCGACATCCGCCACGCCGCCCAGCAGCCGTTCGACCTGGCCG 43981

Qy        484 ATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGA 543
              |  | | | ||| | |||   |    ||| |    ||  |  ||   |  || | ||
Db      43980 ACGTGGATCCCGCGCGCCTGCGGGAGATCCTCGAACTGCTCCTCGACCTGTTCGACCGCG 43921

Qy        544 CCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCA 603
              || | |  |  | || || ||| |     ||||||   ||  ||    ||  | ||
Db      43920 GTGAGCTGTGCCCGCTGCCGGTGCAGCCGTGGGACATCCGCCGTGCGCGGGACGCGTTCG 43861

Qy        604 GCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCG 663
              |||  || || | ||  |  |||| | |  |   |  |  |    |  |||
Db      43860 CGTGGATGAGCCACGCCCGGCACACCGGCAAGATGGTCCTCACCGTGCCGCCTCCGATCG 43801

Qy        664 GGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCA 723
              |  ||      ||  |||  | |  |||  | || ||   | |||| | |   | |
Db      43800 GCCCGGACGCGCCGGTCCTGGTCACCGGGGAGGACGCCGACACCGTCGCCGGCCTGCTCC 43741
```

```
Qy        724 CCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCG 783
              |     |  ||  |||  |   |     |         ||| |  | ||   |||||| ||
Db      43740 GCGACGAGACCGGCCTCACCCACGTGACCACCGCAGCCGAGGAAGCCGCGAACCCCGGCG 43681

Qy        784 AGCTCCGCCTCGTCCTGCACTGGGGCGCGG 813
              ||||   | |||||||| |   ||| |||
Db      43680 CGCTCGTGCACGTCCTGCTCGACGGCCCGG 43651


RESULT 6
ADQ91713
ID   ADQ91713 standard; DNA; 22533 BP.
XX
AC   ADQ91713;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Polyketide synthase ORF17, SEQ ID 36.
XX
KW   Antifungal; Cytostatic; polyketide synthase; enzyme; polyene polyketide;
KW   polyketide; cancer; gene; ds.
XX
OS   Streptomyces aizunensis; NRRL B-11277.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .22533
FT                   /*tag=  a
FT                   /product= "Polyketide synthase ORF17 protein"
XX
PN   WO2004065401-A1.
XX
PD   05-AUG-2004.
XX
PF   21-JAN-2004; 2004WO-CA000068.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Bachmann BO,  Mcalpine JB,  Zazopoulos E,  Farnet CM;
XX
DR   WPI; 2004-348649/33.
DR   P-PSDB; ADQ91712.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Disclosure; SEQ ID NO 36; 554pp; English.
XX
CC   The present invention relates to novel Streptomyces aizunensis (NRRL B-
CC   11277) polyketide synthases and their coding sequences for polyene
CC   polyketide production. The polyketides are useful as fungal growth
CC   inhibitors and cancer cell growth inhibitors. The present sequence is a
CC   coding sequence for a polyketide synthase of the invention.
XX
SQ   Sequence 22533 BP; 2786 A; 8775 C; 7780 G; 3192 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 87.8;  DB 13;  Length 22533;
  Best Local Similarity   44.4%;  Pred. No. 0.00011;
  Matches  399;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||   |||||  |    | | |    |  |||||||| | |||  | ||    |||   |
Db      10554 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 10613

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              |||   |||||    |   | | ||| |||| ||||   |||| ||  |  ||
Db      10614 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 10673

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              || ||   || |  |   ||  | | |  | ||    ||     |  || |  | |
Db      10674 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 10733

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              || || |  ||||| |   | || |  ||  | || |      |  |||    | |
Db      10734 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 10793

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              || ||   |  |  |   ||   ||  |         | ||    || || || ||
Db      10794 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 10853

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
              || ||||  |  || |  ||  ||       |     | | |  |  |  |  | |
Db      10854 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 10913

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              || |  |  |   |  ||||| || || || |||      |  | ||| ||| | | |
Db      10914 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 10973

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              | | | |   || |  | | |  | | |  ||  | ||| | || |||
Db      10974 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 11033

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||| ||   |  |  | ||||     || | | ||||          |  |  || |  |
Db      11034 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 11093

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | || |    |  |  | |    | ||||   || | |||  ||       |||
Db      11094 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 11153

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||     || |  | | |    | ||||       |  | || || ||  |
Db      11154 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 11213

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||         | | ||  |      |  ||   ||  |||   |||||   |||   |
Db      11214 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 11273

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
              |      |      ||   ||||||| | || ||   |   | |||||||  ||  || |||
Db      11274 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 11333

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
              |   | |||  ||| |  | ||   |  |  |  |     |  |  |  | || | || |
Db      11334 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 11393

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
              || |   |||||  |  | |  |||  |   |   ||   |  |   ||||||
Db      11394 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 11452


RESULT 7
AEB87005
ID   AEB87005 standard; DNA; 22533 BP.
XX
AC   AEB87005;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Streptomyces aizunensis polyketide synthase ORF17 DNA SEQ ID NO:36.
XX
KW   polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW   gene; ds.
XX
OS   Streptomyces aizunensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .22533
FT                   /*tag=  a
FT                   /product= "polyene polyketide synthase ORF17"
XX
```

```
PN   CA2453071-A1.
XX
PD   03-APR-2004.
XX
PF   21-JAN-2004; 2004CA-02453071.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazapoulos E;
XX
DR   WPI; 2004-348649/33.
DR   P-PSDB; AEB87004.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Claim 2; SEQ ID NO 36; 550pp; English.
XX
CC   The invention relates to an isolated, purified or enriched nucleic acid
CC   which can be used for the production of a polyketide. The nucleic acid
CC   encodes at least one domain of the polyketide synthase system formed by
CC   the polyketide synthases of AEB86865, AEB86867, AEB86869, AEB86871,
CC   AEB86873, AEB86875, AEB86877, AEB86879 and AEB86881. Also described: (1)
CC   an isolated, purified or enriched nucleic acid for the production of a
CC   polyketide of Formula I comprising a nucleic acid selected from the group
CC   consisting of AEB86866, AEB86868, AEB86870, AEB86872, AEB86874, AEB86876,
CC   AEB86878, AEB86880 and AEB86882; (2) an isolated, purified or enriched
CC   nucleic acid for the production of a polyene polyketide; (3) an
CC   expression vector comprising the novel nucleic acid or the nucleic acid
CC   of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC   the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC   250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC   the production of a polyketide. The nucleic acid can be used for
CC   production of a polyketide, for use as a fungal and cancer cell growth
CC   inhibitors. The present sequence encodes Streptomyces aizunensis polyene
CC   polyketide synthase ORF17, which is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 22533 BP; 2786 A; 8775 C; 7780 G; 3192 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 87.8;  DB 13;  Length 22533;
  Best Local Similarity   44.4%;  Pred. No. 0.00011;
  Matches  399;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||   |||||  |    | | |    |  |||||||| | |||  | ||    |||   |
Db      10554 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 10613

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              |||   |||||    |   | | ||| |||| ||||   |||| ||  |  ||
Db      10614 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 10673

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              || ||    || |  |    ||   | | |  | ||     ||     |  || |  | |
Db      10674 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 10733

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
                || | |||||  |    | || |  ||  | || |      |  |||    | |
Db      10734 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 10793

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              || ||   |  |  |   ||  ||  |         |  ||    || || || || ||
Db      10794 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 10853

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
                || ||||  |  || |  ||  ||     |       |  |  |  |  |  | | |
Db      10854 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 10913

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              || |  |  |   |    ||||| || || || |||       |  | ||| ||| | | |
Db      10914 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 10973

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              |  | |  |   || |   | | |  |  | |   ||  | ||| | || |||
Db      10974 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 11033

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||| ||   |  |  | ||||      || | | ||||        |  |  || | |
Db      11034 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 11093

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | || |   |  |  | |   |  ||||  || | |||  ||      |||
Db      11094 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 11153

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||     || |  | | |    |  ||||          |  | || || ||  |
Db      11154 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 11213

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||        | | ||  |        |  ||   ||  |||   |||||   |||   |
Db      11214 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 11273

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
              |    |    ||   |||||| | || ||  |   | |||||||  ||  || |||
Db      11274 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 11333

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
              |   | |||  ||| |  | ||   |  | | |     | | |  | ||| | || | ||
Db      11334 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 11393

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
              || |   ||||| |   | |  ||| |  |  | |   ||   |  |   ||||||
Db      11394 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 11452


RESULT 8
AEB86802
ID   AEB86802 standard; DNA; 22533 BP.
XX
AC   AEB86802;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Streptomyces aizunensis polyketide synthase ORF17 DNA SEQ ID NO:36.
XX
KW   polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW   gene; ds.
XX
OS   Streptomyces aizunensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .22533
FT                   /*tag=  a
FT                   /product= "polyene polyketide synthase ORF17"
XX
PN   CA2453080-A1.
XX
PD   03-APR-2004.
XX
PF   21-JAN-2004; 2004CA-02453080.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Bachmann BO,  Mcalpine JB,  Zazopoulos E;
XX
DR   WPI; 2004-348650/33.
```

```
DR   P-PSDB; AEB86801.
XX
PT   Polyene polyketide compounds are bacterial cell growth inhibitors useful
PT   for treating e.g. fungal infections and cancer.
XX
PS   Claim 2; SEQ ID NO 36; 144pp; English.
XX
CC   The invention relates to an isolated, purified or enriched nucleic acid
CC   which can be used for the production of a polyketide. The nucleic acid
CC   encodes at least one domain of the polyketide synthase system formed by
CC   the polyketide synthases of AEB86662, AEB86664, AEB86666, AEB86668,
CC   AEB86670, AEB86672, AEB86674, AEB86676 and AEB86678. Also described: (1)
CC   an isolated, purified or enriched nucleic acid for the production of a
CC   polyketide of Formula I comprising a nucleic acid selected from the group
CC   consisting of AEB86663, AEB86665, AEB86667, AEB86669, AEB86671, AEB86673,
CC   AEB86675, AEB86677 and AEB86679; (2) an isolated, purified or enriched
CC   nucleic acid for the production of a polyene polyketide; (3) an
CC   expression vector comprising the novel nucleic acid or the nucleic acid
CC   of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC   the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC   250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC   the production of a polyketide. The nucleic acid can be used for
CC   production of a polyketide, for use as a fungal and cancer cell growth
CC   inhibitors. The present sequence encodes Streptomyces aizunensis polyene
CC   polyketide synthase ORF17, which is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 22533 BP; 2786 A; 8775 C; 7780 G; 3192 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 87.8;  DB 13;  Length 22533;
  Best Local Similarity   44.4%;  Pred. No. 0.00011;
  Matches  399;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||   |||||  |    | | |    | ||||||| | |||  | ||    |||   |
Db      10554 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 10613

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              |||   |||||    |   | | ||| |||| ||||   |||| ||  |  ||
Db      10614 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 10673

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              || ||    || |  |    ||   | | |  | ||     ||     |  || |  | |
Db      10674 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 10733

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              || || |  |||||  |   | ||| |  ||  | || |     |  |||    | |
Db      10734 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 10793

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              || ||   |  |  |   ||   ||  |      |  ||   || || || || ||
Db      10794 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 10853

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
              || ||||  |  || |   ||  ||     |       |  |  |  |  |  | | |
Db      10854 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 10913

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              || |  |  |    |    ||||| || || ||| |   | |||| ||| | | |
Db      10914 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 10973

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              |  | |  |   || |  | | |  |  ||  | ||| | || |||
Db      10974 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 11033

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||| ||   |  |  | ||||      || | | ||||       |  |  || |  |
Db      11034 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 11093

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | || |    |  | | |    | ||||  || | |||  ||      |||
Db      11094 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 11153

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||      || |  | | |    |  ||||           |  | || || ||  |
Db      11154 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 11213

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||         | | ||  |         |  ||   ||  |||   |||||  |||  |
Db      11214 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 11273

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
              |    |      ||   ||||||| | || ||  |   | |||||||  ||  || |||
Db      11274 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 11333

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
              |    | |||  ||| | | ||   | |  | |      |  |  |  | ||| | || |
Db      11334 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 11393

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
                || |   ||||| |   | |  ||| |  |  |  |   ||   |  |  ||||||
Db      11394 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 11452


RESULT 9
ADQ91695
ID   ADQ91695 standard; DNA; 164051 BP.
XX
AC   ADQ91695;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Polyketide synthase related DNA contig 2, SEQ ID 18.
XX
KW   Antifungal; Cytostatic; polyketide synthase; enzyme; polyene polyketide;
KW   polyketide; cancer; gene; ds.
XX
OS   Streptomyces aizunensis; NRRL B-11277.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(6. .1109)
FT                   /*tag=  a
FT                   /product= "Polyketide synthase related ORF9 protein"
FT   CDS             1375. .25818
FT                   /*tag=  b
FT                   /product= "Polyketide synthase ORF10 protein"
FT   CDS             25902. .36188
FT                   /*tag=  c
FT                   /product= "Polyketide synthase ORF11 protein"
FT   CDS             36213. .56468
FT                   /*tag=  d
FT                   /product= "Polyketide synthase ORF12 protein"
FT   CDS             56600. .61573
FT                   /*tag=  e
FT                   /product= "Polyketide synthase ORF13 protein"
FT   CDS             61852. .77475
FT                   /*tag=  f
FT                   /product= "Polyketide synthase ORF14 protein"
FT   CDS             77606. .93904
FT                   /*tag=  g
FT                   /product= "Polyketide synthase ORF15 protein"
FT   CDS             94057. .103740
FT                   /*tag=  h
FT                   /product= "Polyketide synthase ORF16 protein"
FT   CDS             103789. .126321
FT                   /*tag=  i
FT                   /product= "Polyketide synthase ORF17 protein"
FT   CDS             126389. .138007
FT                   /*tag=  j
FT                   /product= "Polyketide synthase ORF18 protein"
FT   CDS             complement(138063. .139079)
FT                   /*tag=  k
FT                   /product= "Polyketide synthase related ORF19 protein"
FT   CDS             complement(139266. .140117)
FT                   /*tag=  l
FT                   /product= "Polyketide synthase related ORF20 protein"
```

```
FT   CDS             complement(140114. .141103)
FT                   /*tag= m
FT                   /product= "Polyketide synthase related ORF21 protein"
FT   CDS             141483. .142436
FT                   /*tag= n
FT                   /product= "Polyketide synthase related ORF22 protein"
FT   CDS             142440. .143054
FT                   /*tag= o
FT                   /product= "Polyketide synthase related ORF23 protein"
FT   CDS             143133. .144119
FT                   /*tag= p
FT                   /product= "Polyketide synthase related ORF24 protein"
FT   CDS             144116. .145102
FT                   /*tag= q
FT                   /product= "Polyketide synthase related ORF25 protein"
FT   CDS             145099. .145743
FT                   /*tag= r
FT                   /product= "Polyketide synthase related ORF26 protein"
FT   CDS             145818. .147230
FT                   /*tag= s
FT                   /product= "Polyketide synthase related ORF27 protein"
FT   CDS             complement(147306. .148967)
FT                   /*tag= t
FT                   /product= "Polyketide synthase related ORF28 protein"
FT   CDS             complement(149176. .149871)
FT                   /*tag= u
FT                   /product= "Polyketide synthase related ORF29 protein"
FT   CDS             complement(149868. .150788)
FT                   /*tag= v
FT                   /product= "Polyketide synthase related ORF30 protein"
FT   CDS             complement(150769. .153765)
FT                   /*tag= w
FT                   /product= "Polyketide synthase related ORF31 protein"
FT   CDS             154485. .156041
FT                   /*tag= x
FT                   /product= "Polyketide synthase related ORF32 protein"
FT   CDS             156075. .157064
FT                   /*tag= y
FT                   /product= "Polyketide synthase related ORF33 protein"
FT   CDS             157308. .158873
FT                   /*tag= z
FT                   /product= "Polyketide synthase related ORF34 protein"
FT   CDS             158970. .160202
FT                   /*tag= aa
FT                   /product= "Polyketide synthase related ORF35 protein"
FT   CDS             160199. .161719
FT                   /*tag= ab
FT                   /product= "Polyketide synthase related ORF36 protein"
FT   CDS             161924. .162577
FT                   /*tag= ac
FT                   /product= "Polyketide synthase related ORF37 protein"
FT   CDS             162723. .164051
FT                   /*tag= ad
FT                   /product= "Polyketide synthase related ORF38 protein"
XX
PN   WO2004065401-A1.
XX
PD   05-AUG-2004.
XX
PF   21-JAN-2004; 2004WO-CA000068.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Bachmann BO,  Mcalpine JB,  Zazopoulos E,  Farnet CM;
XX
DR   WPI; 2004-348649/33.
DR   P-PSDB; ADQ91696, ADQ91698, ADQ91700, ADQ91702, ADQ91704, ADQ91706,
DR   ADQ91708, ADQ91710, ADQ91712, ADQ91714, ADQ91716, ADQ91718, ADQ91720,
DR   ADQ91722, ADQ91724, ADQ91726, ADQ91728, ADQ91730, ADQ91732, ADQ91734,
DR   ADQ91736, ADQ91738, ADQ91740, ADQ91742, ADQ91744, ADQ91746, ADQ91748,
DR   ADQ91750, ADQ91752, ADQ91754.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Disclosure; SEQ ID NO 18; 554pp; English.
XX
CC   The present invention relates to novel Streptomyces aizunensis (NRRL B-
CC   11277) polyketide synthases and their coding sequences for polyene
CC   polyketide production. The polyketides are useful as fungal growth
CC   inhibitors and cancer cell growth inhibitors. The present sequence is a
CC   DNA contig which comprises thirty Open Reading Frames (ORFs), encoding
CC   polyketide synthases and related proteins.
XX
SQ   Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 87.8;  DB 13;  Length 164051;
  Best Local Similarity   44.4%;  Pred. No. 8.4e-05;
  Matches  399;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||  |||||  |    | | |    | ||||||| | |||  | ||   |||   |
Db     114342 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 114401

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              |||   |||||    |   | | ||| |||| ||||   |||| ||  |  ||
Db     114402 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 114461

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              || ||    || |  |   ||  | | |  | ||    ||    |  || |  | |
Db     114462 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 114521

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              || || |  |||||  |   | || |  ||  | || |     |  |||    | |
Db     114522 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 114581

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              || ||   |  |  |   ||   ||  |         |  ||    || || || || ||
Db     114582 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 114641

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
              || ||||  |  || |  ||  ||        |  |  |  |  |  | | |
Db     114642 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 114701

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              || |  |  |  |  ||||| || || || |||     |  | ||| ||| | | |
Db     114702 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 114761

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              | | | |  || |  | | |  | | |  || | ||| | || |||
Db     114762 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 114821

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||| ||  |  |  | ||||     || | | |||       |  |  || |  |
Db     114822 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 114881

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | || |   |  | | |  |  ||||  || | |||  ||      |||
Db     114882 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 114941

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||    || |  | | |    |  ||||          |  | || || ||  |
Db     114942 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 115001

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||       | | ||  |    |  ||   ||  |||   |||||   |||  |
Db     115002 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 115061

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
              |      |      ||   ||||||| | || ||   |  | |||||||  ||  || |||
Db     115062 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 115121
```

```
Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
              |    | |||  ||| |  | ||   |  |  | |       |  |  |  | || | || |
Db     115122 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 115181

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
                || |  |||| |   | |  ||| |  |  |  ||   |  |  |||||
Db     115182 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 115240


RESULT 10
AEB86862
ID   AEB86862 standard; DNA; 164051 BP.
XX
AC   AEB86862;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Streptomyces aizunensis polyketide synthase contig 2 SEQ ID NO:18.
XX
KW   polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW   ds.
XX
OS   Streptomyces aizunensis.
XX
PN   CA2453071-A1.
XX
PD   03-APR-2004.
XX
PF   21-JAN-2004; 2004CA-02453071.
XX
PR   21-JAN-2003; 2003US-0441123P.
PR   13-MAY-2003; 2003US-0469810P.
PR   01-AUG-2003; 2003US-0491516P.
PR   13-AUG-2003; 2003US-0494568P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazapoulos E;
XX
DR   WPI; 2004-348649/33.
XX
PT   Isolated, purified or enriched nucleic acid useful for producing
PT   polyketides, encodes domains of the polyketide synthase system.
XX
PS   Disclosure; SEQ ID NO 18; 550pp; English.
XX
CC   The invention relates to an isolated, purified or enriched nucleic acid
CC   which can be used for the production of a polyketide. The nucleic acid
CC   encodes at least one domain of the polyketide synthase system formed by
CC   the polyketide synthases of AEB86865, AEB86867, AEB86869, AEB86871,
CC   AEB86873, AEB86875, AEB86877, AEB86879 and AEB86881. Also described: (1)
CC   an isolated, purified or enriched nucleic acid for the production of a
CC   polyketide of Formula I comprising a nucleic acid selected from the group
CC   consisting of AEB86866, AEB86868, AEB86870, AEB86872, AEB86874, AEB86876,
CC   AEB86878, AEB86880 and AEB86882; (2) an isolated, purified or enriched
CC   nucleic acid for the production of a polyene polyketide; (3) an
CC   expression vector comprising the novel nucleic acid or the nucleic acid
CC   of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC   the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC   250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC   the production of a polyketide. The nucleic acid can be used for
CC   production of a polyketide, for use as a fungal and cancer cell growth
CC   inhibitors. The present sequence represents a Streptomyces aizunensis
CC   polyene polyketide synthase related contig 2 DNA sequence, which is used
CC   in the exemplification of the present invention.
XX
SQ   Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 87.8;  DB 13;  Length 164051;
  Best Local Similarity   44.4%;  Pred. No. 8.4e-05;
  Matches  399;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||   |||||  |    | | |    | |||||||| | |||  | ||   |||  |
Db     114342 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 114401

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              |||  |||||    |  | | ||| |||| ||||   |||| ||  |  ||
Db     114402 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 114461

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              || ||   || |  |    ||   | |  | ||     ||     |  || |  | |
Db     114462 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 114521

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              ||  || |  ||||| |   | || |  ||  | || |       |  |||    | |
Db     114522 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 114581

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              || ||   |  |  |  ||  ||  |          |  ||    || || || || ||
Db     114582 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 114641

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
                || ||||  |  || |  ||  ||     |     |  |  |  |  |  | | |
Db     114642 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 114701

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              || |  | |  |  |  ||||| || || || |||       |  | ||| ||| | | |
Db     114702 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 114761

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              | | | |   || |  | | |  | | |  ||  | ||| | || |||
Db     114762 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 114821

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||| ||   |  |  | ||||     || | | ||||       |  |  || |  |
Db     114822 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 114881

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | || |    |  |  | |    | ||||  || | |||  ||       |||
Db     114882 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 114941

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||     || |  | | |    |  ||||           |  | || || ||  |
Db     114942 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 115001

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||       | | ||  |      |  ||  ||  |||   |||||  |||  |
Db     115002 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 115061

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
              |     |      ||  |||||||  | || ||  |  |||||||  ||  || |||
Db     115062 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 115121

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
              |    | |||  ||| |  | ||   |  |  | |       |  |  |  | || | || |
Db     115122 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 115181

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
                || |  |||| |   | |  ||| |  |  |  ||   |  |  |||||
Db     115182 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 115240


RESULT 11
AEB86659
ID   AEB86659 standard; DNA; 164051 BP.
XX
AC   AEB86659;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Streptomyces aizunensis polyketide synthase contig 2 SEQ ID NO:18.
XX
KW   polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;
KW   ds.
```

XX
OS Streptomyces aizunensis.
XX
PN CA2453080-A1.
XX
PD 03-APR-2004.
XX
PF 21-JAN-2004; 2004CA-02453080.
XX
PR 21-JAN-2003; 2003US-0441123P.
PR 13-MAY-2003; 2003US-0469810P.
PR 01-AUG-2003; 2003US-0491516P.
PR 13-AUG-2003; 2003US-0494568P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Bachmann BO, Mcalpine JB, Zazopoulos E;
XX
DR WPI; 2004-348650/33.
XX
PT Polyene polyketide compounds are bacterial cell growth inhibitors useful
PT for treating e.g. fungal infections and cancer.
XX
PS Disclosure; SEQ ID NO 18; 144pp; English.
XX
CC The invention relates to an isolated, purified or enriched nucleic acid
CC which can be used for the production of a polyketide. The nucleic acid
CC encodes at least one domain of the polyketide synthase system formed by
CC the polyketide synthases of AEB86662, AEB86664, AEB86666, AEB86668,
CC AEB86670, AEB86672, AEB86674, AEB86676 and AEB86678. Also described: (1)
CC an isolated, purified or enriched nucleic acid for the production of a
CC polyketide of Formula I comprising a nucleic acid selected from the group
CC consisting of AEB86663, AEB86665, AEB86667, AEB86669, AEB86671, AEB86673,
CC AEB86675, AEB86677 and AEB86679; (2) an isolated, purified or enriched
CC nucleic acid for the production of a polyene polyketide; (3) an
CC expression vector comprising the novel nucleic acid or the nucleic acid
CC of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC the cosmid deposited under IDAC accession no. 250203-01, 250203-02,
CC 250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC the production of a polyketide. The nucleic acid can be used for
CC production of a polyketide, for use as a fungal and cancer cell growth
CC inhibitors. The present sequence represents a Streptomyces aizunensis
CC polyene polyketide synthase related contig 2 DNA sequence, which is used
CC in the exemplification of the present invention.
XX
SQ Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;

Query Match 9.7%; Score 87.8; DB 13; Length 164051;
Best Local Similarity 44.4%; Pred. No. 8.4e-05;
Matches 399; Conservative 0; Mismatches 497; Indels 3; Gaps 1;

```
Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||   |||||  |    | | |    |  |||||||| | |||  | ||    |||   |
Db     114342 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 114401

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              |||   |||||    |   | | ||| |||| ||||   |||| ||  |  ||
Db     114402 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 114461

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              || ||    || |  |    ||   | | |  | ||     ||      |  || |  | |
Db     114462 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 114521

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              || || |  |||||  |   | || |  ||  | || |    |  |||    | |
Db     114522 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 114581

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              || ||   |  |  |   ||   ||  |        |  ||    || || || || ||
Db     114582 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 114641

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
                || ||||  |  || |   ||  ||     |       |  |  |  |  |  | |
Db     114642 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 114701

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              || |  |  |   |    ||||| || || || |||      |  | ||| ||| | | |
Db     114702 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 114761

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              | | |  |   || |  | | |  | | |   || | ||| | || |||
Db     114762 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 114821

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||| ||   |  |  | ||||      || | | ||||        |  |  || |  |
Db     114822 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 114881

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | || |    |  |  | |    | ||||   || | |||  ||       |||
Db     114882 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 114941

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||     || |  | | |    | ||||      |  | || || ||  |
Db     114942 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 115001

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||       | | ||  |      |  ||   ||  |||   |||||   |||   |
Db     115002 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 115061

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
              |    |      ||   ||||||| | || ||  |   | |||||||  ||  || |||
Db     115062 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 115121

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
              |   | |||  ||| | | ||   |  | | |  |  |  |  | | | | || | ||
Db     115122 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 115181

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
              || |   ||||| |   | |  ||| |  |  |  |   ||   |  |  ||||||
Db     115182 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 115240
```

RESULT 12
ACA37866
ID ACA37866 standard; DNA; 3948 BP.
XX
AC ACA37866;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #19523.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Mycobacterium avium.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU33996.

```
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 25736; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3948 BP; 724 A; 1294 C; 1357 G; 573 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 87.4;  DB 8;  Length 3948;
  Best Local Similarity   45.9%;  Pred. No. 0.00017;
  Matches  390;  Conservative    0;  Mismatches  441;  Indels   18;  Gaps    2;

Qy         43 TCTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGC 102
              ||| ||| || ||    ||| ||||| | ||| | |    |     |||||| | ||  |
Db        221 TCTGTGAGCGCTGCGGCGTCGAGGTGACCCGCGCCAAGGTGCGCCGCGAGCGCATGGGCC 280

Qy        103 AGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGC 162
              |   |          ||| || | ||  | || | |    | | |   |||||
Db        281 ACATCGAGTTGGCCGCGCCCGTCACGCACATCTGGTACTTCAAGGGCGTGCCGTCGCGAT 340

Qy        163 CCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCC 222
                  ||||| ||| ||| | | | |  ||  |||    ||    ||  |  || ||
Db        341 TGGGCTACCTGCTCGACCTGGCGCCCAAGGACCTCGAGAAGATCATCTACTTCGCCGCGT 400

Qy        223 GGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGG 282
                |   | ||     | |||  |  ||  | |||   |||| ||   | | | |    |
Db        401 ACGTGATCACGTCCGTCGACGAGGAGATGCGGCACAACGAGCTCTCCACCCTCGAGGCCG 460

Qy        283 ATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCG 342
              |  | | || ||   | ||  ||        | |||||   | || |   |||| |  ||
Db        461 AGATGATGGTGGAGCGCAAGGCCGTCGAGGACCAGCGCGACGCCGACCTGGAGGCGCGCG 520

Qy        343 GCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCG 402
                | ||  | ||    ||  ||| | | | || | |||||| | |||| | | ||| |
Db        521 CCCAGAAGCTGGAGGCCGACCTGGCCGAGCTGGAGGCCGAGGGCGCCAAGGCCGACGCGC 580

Qy        403 AGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACC 462
                | ||  |  ||         ||    |        ||    ||||  | ||||||| |
Db        581 GGCGCAAGGTCCGCGACAGCGGCGAGCGGGAGATGCGCCAGATCCGCGACCGCGCCCAGC 640

Qy        463 TCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGC 522
              ||| | |  |||   ||| ||  ||      | || ||  | ||||   |||      ||
Db        641 GCGAGCTGGACCGGCTGGAGGACATCTGGAACACCTTCACCAAGCTGGCCCCCAAGCAGC 700

Qy        523 GGCGGGAGGAGGCG----------GAGCCGACCGATCGGGTCGAGGTGGCGATGCTCC 570
              |   | || | |            |||| |  ||| ||   ||  | |    |   |
Db        701 TGATCGTCGACGAGAACCTCTACCGCGAGCTGGTCGACCGTTACGGCGAGTACTTCACCG 760

Qy        571 TCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCG 630
              | |    || |  ||||| ||| | | ||| | ||| |  |||||    |  ||  |
Db        761 GCGCCATGGGCGCGGAGTCGATCCAGAAGCTCATCGAGAACTTCGACATCGACGCCGAGG 820

Qy        631 GCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGG 690
              | | |  | |||   |  |  |     |  |  |  |  |||    |  | |
Db        821 CCGAGCAGCTGCGTGACGTCATCCGAAACGGCAAGGGGCAGAAGAAGCTTCGCGCCCTCA 880

Qy        691 GGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCC 750
              |     | | |    | |||| || | |  | |    |      || |     || |
Db        881 AGCGGCTCAAGGTGGTCGCCGCGTTCCAACAGAGCGGCAACTCGCCGAT------GGGCA 934

Qy        751 GCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCG 810
              ||||||  || ||     |    |||  | ||||| ||||   |  ||||    | ||
Db        935 TGGTGCTCGACGCCGTCCCGGTGATCCCGCCGGAGCTGCGCCCGATGGTGCAGCTCGACG 994

Qy        811 CGGTGCGGATGAGCGATCACGACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCG 870
              | |||  |    |      |||| | | || |  | |    ||     | ||  ||| | |
Db        995 GTGGCCGGTTCGCCACCAGCGACCTCAACGACCTGTACCGCCGGGTGATCAACCGCAACA 1054

Qy        871 AGGCGGTCA 879
              |   | |||
Db       1055 ACCGGCTCA 1063


RESULT 13
ADX32903
ID   ADX32903 standard; cDNA; 1254 BP.
XX
AC   ADX32903;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polynucleotide seqid 15723.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content; gene; ss.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
```

```
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 15723; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC   of the invention are also useful in physical arrays of molecules and as
CC   plant breeding markers. The recombinant DNA construct is useful for
CC   improving plant tolerance to cold, heat, drought, herbicides, extreme
CC   osmotic conditions, pathogens or pests, for manipulating growth rate in
CC   plant cells by modification of the cell cycle pathway, for conferring
CC   increased resistance to plant disease, for producing galactomannan,
CC   lignin or plant growth regulators, for increasing the rate of homologous
CC   recombination in plants, for improving yield by modification of
CC   photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC   or by providing improved plant growth and development under at least one
CC   stress condition or for modifying seed oil or protein yield and/or
CC   content. This sequence represents a plant full length insert
CC   polynucleotide that can be used in the recombinant DNA construct of the
CC   invention.
XX
SQ   Sequence 1254 BP; 192 A; 466 C; 408 G; 188 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 87;  DB 13;  Length 1254;
  Best Local Similarity   46.6%;  Pred. No. 0.00024;
  Matches  405;  Conservative    0;  Mismatches  440;  Indels   24;  Gaps    3;

Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
              ||   | | |  |||||  | | || |  ||  |||||        ||  |  | |
Db        279 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 338

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
               |   |  | |||     || |  | || |  || |  || || |  | | | ||  | |
Db        339 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 398

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
              ||  ||| |  | |||   ||   |||||||| ||  ||||| | |  |  | |||    |
Db        399 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 455

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              || | |   |   | ||||    ||      |   |||  |||  |    || || |  |
Db        456 CTGCTGAGCCAGCTGGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 515

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
              | ||| |  ||  |   || | |||| ||          | ||||   || |  ||||
Db        516 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 575

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
              | |||  |     | |   || | || | || ||| ||| ||  |               |
Db        576 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC------------C 623

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              || ||| ||||  |  | |||    || ||| | || |  |||||||||   |||| ||
Db        624 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 683

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
                  | ||||  || |  |  | || |  ||  || | |  |   ||  ||||
Db        684 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 734

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              ||  ||| ||    |||| |  | |  | |||| |  |  || |   |  |     |  |
Db        735 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 794

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
                ||   | |||||    | |     | ||||||   | |||         ||
Db        795 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 854

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
                ||    || |      | | |  | |        || | | || || |   | |
Db        855 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 914

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
              |   | ||| ||      | ||| |||| |  ||    | |||||  | | ||   |
Db        915 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCGCGCAAGCCACTC 974

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
              |      ||      |||| | |  || | |  | ||||   ||   ||  ||||| |
Db        975 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCACACGCGCACGGCGCCGTG 1034

Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845
              ||| |||     |   | |||  ||| | ||  |    | | |||| |   ||| |    |
Db       1035 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1094

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874
              | || |     ||   | | |  | ||||
Db       1095 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1123


RESULT 14
ADX32964
ID   ADX32964 standard; cDNA; 1434 BP.
XX
AC   ADX32964;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polynucleotide seqid 15784.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content; gene; ss.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 15784; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
```

CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 1434 BP; 256 A; 486 C; 457 G; 235 T; 0 U; 0 Other;

Query Match 9.7%; Score 87; DB 13; Length 1434;
Best Local Similarity 46.6%; Pred. No. 0.00023;
Matches 405; Conservative 0; Mismatches 440; Indels 24; Gaps 3;

```
Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
              ||   | | |  |||||  | | || |  ||  |||||            ||  |  |
Db        269 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 328

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
                |   | | |||     || |  | || |  || |  || || |  | | | ||  | |
Db        329 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 388

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
              ||  ||| |  | |||   ||   ||||||| ||  ||||| | |  |  | |||    |
Db        389 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 445

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              || | |   |   | ||||    ||      |   |||  |||  |    || || |  |
Db        446 CTGCTGAGCCAGCTGGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 505

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
                | ||| |  ||  |   || | |||| ||          | ||||   || |  ||||
Db        506 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 565

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
              | |||  |     | |   || | || | || ||| ||| ||  |             |
Db        566 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC----------C 613

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              || ||| |||   | | |||    || ||| | || |  |||||||||    |||| ||
Db        614 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 673

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
                  | ||||  || |  |  |  || |  ||  || | |  |  ||  ||||
Db        674 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 724

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              ||  ||| ||    |||| |  | |  | |||| |  |  || |   |  |    |  |
Db        725 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 784

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
                 ||     | ||||      |  |       | ||||||     | |||       ||
Db        785 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 844

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
                ||    || |     | | |  | |     || | | || || |   | |
Db        845 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 904

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
              |   | ||| ||       | ||| ||||  |  ||    | |||||  | | ||   |
Db        905 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCGCGCAAGCCACTC 964

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
              |    ||      |||| | |  || | |  |  ||||   ||   ||  |||||  |
Db        965 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCACACGCGCACGGCGCCGTG 1024
```

```
Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845
              ||| |||     |   | |||  ||| | ||  |    | | ||||  |   ||| |    |
Db       1025 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1084

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874
              | || |     ||   | | |  | ||||
Db       1085 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1113
```

RESULT 15
AAL61171
ID AAL61171 standard; DNA; 9222 BP.
XX
AC AAL61171;
XX
DT 22-SEP-2003 (first entry)
XX
DE Actinosynnema pretiosum polyketide synthase (PKS) gene #2.
XX
KW Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;
KW gene; ds.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
DR WPI; 2003-493374/46.
XX
PT Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS Disclosure; Page 53-58; 160pp; English.
XX
CC The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumour agent. The present sequence is A. pretiosum
CC ansamitocin gene cluster I polyketide synthase (PKS) gene
XX
SQ Sequence 9222 BP; 784 A; 3437 C; 3858 G; 1143 T; 0 U; 0 Other;

Query Match 9.7%; Score 87; DB 8; Length 9222;
Best Local Similarity 44.5%; Pred. No. 0.00017;
Matches 391; Conservative 0; Mismatches 485; Indels 3; Gaps 1;

```
Qy         11 CGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGAT 70
              ||  | |  |||   |  ||  | | ||  | | ||  |  || | ||  |||||  |
Db       1641 CGATGCCCCCGTGCCTGTGCCCGTGCCCGCCCCCGTCCCGGTCGTCGTGTCCGGGGCCAC 1700

Qy         71 CGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGT 130
              |  || || || |   ||| ||  |  |  | |  |   | || || | |      | |
Db       1701 CACCGCCGGTCTGCGCGCCCAGGCGGCGCGCCTGGCCGGGCACCTGCGGGAGCGGCCCGC 1760

Qy        131 CGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCG 190
              |||||  |     |||     | ||     ||      | || |  ||| |      | ||
Db       1761 GCTCGGCCCCGAGGACGTCGCCCGGCCGCTGCTGCTCTCCCGCGCCCAGCGCGAGCGGCG 1820

Qy        191 GGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCAT 250
              ||  || | |  | | |  | | |  |  | | || ||| | |   |  | |||
Db       1821 CGCCGTCGTGGTCGCCCGCGACCGGGAGGCGCTGCTCGGCGGGCTCGACGCCCTCGCGGG 1880

Qy        251 CGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCT 310
```

```
Db       1881 TGGTGAGGCCGGGCGCCGGCTCGTGAGCGGCGCGGCCGACGTCACCGGGCGGGTGGTCCT 1940

Qy        311 CTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCA 370

Db       1941 GGTGTTCCCCGGTCAGGGCGCGCACTGGACCGGCGTGGCCGAGCGGTTGTGGCGGGAGGC 2000

Qy        371 CCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGC 430

Db       2001 GCCCGCCTTCGCCGACAGCATGGCCCGCTGCGC---CGACGTGCTGCGCGACCTCGCCGG 2057

Qy        431 GTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCA 490

Db       2058 GTGGGAGCTGCGGGAGGTGCTCGTCGACCCGGTCGCGCTGGAGCGCGTCGACGTGCTGCA 2117

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550

Db       2118 ACCGGTCTCGTTCGCGGTCGTCGTGTCGCTCGCGGCGCTCTGGGCGTCGGTCGGCGTGCG 2177

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610

Db       2178 GCCCGACGCGGTCGTCGGGCACTCGCAGGGCGAGGTGGCCGCCGCCCACGTCGCGGGCGC 2237

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGAC 670

Db       2238 GCTCACGCTCGCCGAGGCCGCCCGGATCGTCGTGCTCCGCTCCGCGCTGATCGCCCGCGA 2297

Qy        671 GAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGAC 730

Db       2298 GCTGTCCGGCCGCGGCGCGATGCTCACCGTCGTCGCCGACGTCGAGCGGGTCACCGCGCT 2357

Qy        731 GCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCG 790

Db       2358 GCTCGCCGGGTTCGAGGGGCGGGTGTGCGTGGCGGCCGTCAACGGGCCCGCGTCGGTCAC 2417

Qy        791 CCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGGGAGGT 850

Db       2418 CGTCTCCGGCGAGGACGCGGCGGTGCGCGAGTTCGAGCGGGTCCTGTCCGCCCGGCGGGT 2477

Qy        851 CCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGA 889

Db       2478 GCTGCGGTGGCGGCTGCCCGGCGTCGACTTCGCCGGGCA 2516


Search completed: April 27, 2006, 08:00:13
Job time : 438.473 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 185.89 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_75000_75900
Perfect score: 901
Sequence: 1 acggccgggacgcggacgac.....gaggaggatcgcgccgcgga 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	95.4	10.6	1926	3	US-09-249-585A-4	Sequence 4, Appli
	2	95.4	10.6	1931	2	US-09-130-114-2	Sequence 2, Appli
c	3	88.4	9.8	77536	3	US-09-410-551B-1	Sequence 1, Appli
c	4	88.4	9.8	77536	3	US-09-940-316B-1	Sequence 1, Appli
	5	85.4	9.5	58857	3	US-09-477-962-1	Sequence 1, Appli
	6	83.2	9.2	2312	2	US-07-736-178C-1	Sequence 1, Appli
	7	82.6	9.2	1068	3	US-09-000-064-1	Sequence 1, Appli
	8	81.8	9.1	5802	3	US-10-152-886-84	Sequence 84, Appl
	9	79.4	8.8	1068	3	US-09-000-040-4	Sequence 4, Appli
	10	79	8.8	1239	3	US-09-252-991A-9713	Sequence 9713, Ap
	11	79	8.8	1719	3	US-09-252-991A-9632	Sequence 9632, Ap
c	12	79	8.8	3654	3	US-09-252-991A-9533	Sequence 9533, Ap
	13	78.8	8.7	5877	3	US-10-152-886-54	Sequence 54, Appl
	14	78.2	8.7	44377	2	US-08-804-227C-7	Sequence 7, Appli
	15	78.2	8.7	44377	2	US-08-804-198-1	Sequence 1, Appli
	16	74	8.2	2409	3	US-09-902-540-6842	Sequence 6842, Ap
c	17	74	8.2	3899	3	US-09-902-540-573	Sequence 573, App
c	18	74	8.2	23673	3	US-09-773-816-1	Sequence 1, Appli
	19	73.6	8.2	1074	3	US-09-758-759-170	Sequence 170, App
c	20	73.6	8.2	109519	3	US-09-758-759-1	Sequence 1, Appli
c	21	73.4	8.1	30001	2	US-08-125-468-1	Sequence 1, Appli
c	22	73.4	8.1	30001	2	US-08-474-933-1	Sequence 1, Appli
	23	73.2	8.1	2034	3	US-09-252-991A-11958	Sequence 11958, A
	24	73.2	8.1	2493	3	US-09-252-991A-11987	Sequence 11987, A
c	25	73.2	8.1	2991	3	US-09-252-991A-12025	Sequence 12025, A
	26	73	8.1	5871	3	US-10-152-886-24	Sequence 24, Appl
	27	72.8	8.1	1407	3	US-09-902-540-6215	Sequence 6215, Ap
c	28	72.8	8.1	1434	3	US-09-902-540-4356	Sequence 4356, Ap
	29	72.8	8.1	13332	3	US-09-902-540-1047	Sequence 1047, Ap
c	30	72.8	8.1	23694	3	US-09-902-540-1216	Sequence 1216, Ap
c	31	72.2	8.0	1452	3	US-09-252-991A-3933	Sequence 3933, Ap
	32	72.2	8.0	4284	3	US-09-252-991A-4006	Sequence 4006, Ap
	33	71.6	7.9	1839	3	US-09-758-759-150	Sequence 150, App
	34	71.6	7.9	6854	3	US-09-194-905-7	Sequence 7, Appli
	35	71.6	7.9	6854	3	US-09-922-683-7	Sequence 7, Appli
	36	71.6	7.9	109519	3	US-09-758-759-1	Sequence 1, Appli
	37	71.4	7.9	1288	2	US-08-440-856A-9	Sequence 9, Appli
	38	71.4	7.9	2219	3	US-08-510-646B-17	Sequence 17, Appl
	39	71.4	7.9	53500	3	US-09-266-965-76	Sequence 76, Appl
	40	71	7.9	1056	3	US-09-266-965-66	Sequence 66, Appl
	41	71	7.9	1408	3	US-09-180-109A-5	Sequence 5, Appli
	42	71	7.9	1408	3	US-09-180-109A-7	Sequence 7, Appli
	43	70.4	7.8	11220	3	US-09-105-537-32	Sequence 32, Appl
	44	70.4	7.8	36778	3	US-09-105-537-5	Sequence 5, Appli
	45	70.4	7.8	38506	3	US-09-320-878-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-249-585A-4
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match 10.6%; Score 95.4; DB 3; Length 1926;
Best Local Similarity 46.3%; Pred. No. 1.2e-07;
Matches 352; Conservative 0; Mismatches 406; Indels 3; Gaps 1;

```
Qy      112 GCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACC 171
            |||| |    |  ||   ||||| | | | || ||  ||    | | |   |     ||
Db      258 GCCACCTTGTCCTCGTCCTCGTCCTCGCCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 317

Qy      172 TCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCG 231
            |||||| |    | |||     | | ||   ||  ||| |  || |  ||   ||| ||
Db      318 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCG 377

Qy      232 CGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCG 291
                | | || |||| || ||  || |   |  | |       | || |       || ||
Db      378 TCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 437

Qy      292 TCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGG 351
            ||  |||||    |   |||  |   | ||| | |  || |         || |    | |
Db      438 TCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCG 497

Qy      352 TGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCG 411
            |  || || |||||| || |||| | | |  ||||  |  |  ||    |  | |  ||
Db      498 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCG 557

Qy      412 CGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGA 471
```

```
Db        558 TCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCC--- 614

Qy        472 TCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGG 531
              ||| |     |   ||| || || |||| ||   | | |||  | |   | |  |
Db        615 TCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC 674

Qy        532 AGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGA 591
              | |  |  |||  |  |||     |   |  ||||| || | |||| |   |
Db        675 TCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCG 734

Qy        592 CCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCC 651
              || | |  ||  |    ||  || |     | ||  || || || |   |   ||| ||
Db        735 TCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 794

Qy        652 GCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCG 711
              ||| |||| ||        || || | ||  |    |  |||  |  |     | ||
Db        795 CCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC 854

Qy        712 CGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCG 771
              |   | |  | |  |   |  | || |   | || |    ||| |  ||   |||
Db        855 TCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCTCG 914

Qy        772 ACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACG 831
              |  ||||| |   | |  |||||||||| | |   | |   || |       |     |
Db        915 TCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 974

Qy        832 ACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAG 872
              |    || |    || ||| ||| | | |||||  | ||
Db        975 TCCACCTCCGGCCCCAGCTCCTCCGTCACCTCCGGCCCCAG 1015


RESULT 2
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert A.
;  APPLICANT: Damaj, Bassam B.
;  APPLICANT: Robbins, Alan K.
;  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;  TITLE OF INVENTION: From Multiple Transfected Episomes
;  FILE REFERENCE: 0867/1D903US1
;  CURRENT APPLICATION NUMBER: US/09/130,114
;  CURRENT FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1931
;   TYPE: DNA
;   ORGANISM: EBNA
US-09-130-114-2

  Query Match             10.6%;  Score 95.4;  DB 2;  Length 1931;
  Best Local Similarity   46.3%;  Pred. No. 1.2e-07;
  Matches  352;  Conservative    0;  Mismatches  406;  Indels    3;  Gaps    1;

Qy        112 GCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACC 171
              |||| |    |  ||   |||||  | | | || ||  ||    | | |    |   ||
Db        258 GCCACCTTGTCCTCGTCCTCGTCCTCGCCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 317

Qy        172 TCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCG 231
              ||||||  |     | |||     | | ||   ||  ||| |  ||  |   ||| ||
Db        318 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCG 377

Qy        232 CGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCG 291
              | | || |||| || ||  || |   | | |      | || |       || ||
Db        378 TCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 437

Qy        292 TCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGG 351
              ||  |||||    |  |||  |  | ||| | |  || |       || |     | |
Db        438 TCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCG 497

Qy        352 TGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCG 411
              |  || || |||||| || |||| | | |  ||||  |  |  ||     |  | ||
Db        498 TCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCG 557

Qy        412 CGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGA 471
              |  |   |  ||  | |   | | || ||  |  |||| | ||  |||||  |
Db        558 TCCTCCTCCCCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCC--- 614

Qy        472 TCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGG 531
              ||| |     |   ||| || || |||| ||   | | |||  | |   | |  |
Db        615 TCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCG 674

Qy        532 AGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGA 591
              | |  |  |||  |  |||     |   |  ||||| || | |||| |   |
Db        675 TCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCCCG 734

Qy        592 CCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCC 651
              || | |  ||  |    ||  || |     | ||  || || || |   |   ||| ||
Db        735 TCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCC 794

Qy        652 GCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCG 711
              ||| |||| ||           || || | || |     |  ||| |  |     | ||
Db        795 CCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCC 854

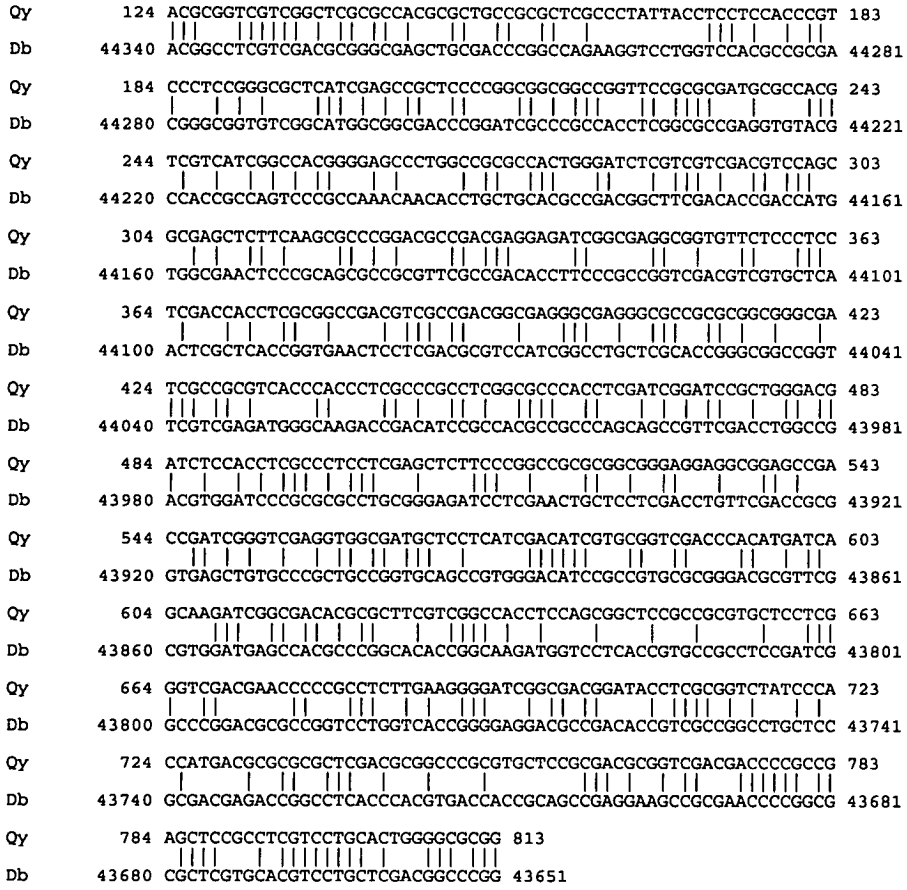
Qy        712 CGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCG 771
              |   | |  | |  |   |  | || |   | || |    ||| |  ||   |||
Db        855 TCCCCGTCCTCGTCCTCCCCGTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCCTCCTCG 914

Qy        772 ACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACG 831
              |  ||||| |   | |  |||||||||| | |   | |   || |       |     |
Db        915 TCCTCCCCGTCCTCCCCGTCCTCGTCCTCCTCCCCGTCCTCGTCCTCCCCGTCCTCGTCC 974

Qy        832 ACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAG 872
              |    || |    || ||| ||| | | |||||  | ||
Db        975 TCCACCTCCGGCCCCAGCTCCTCCGTCACCTCCGGCCCCAG 1015

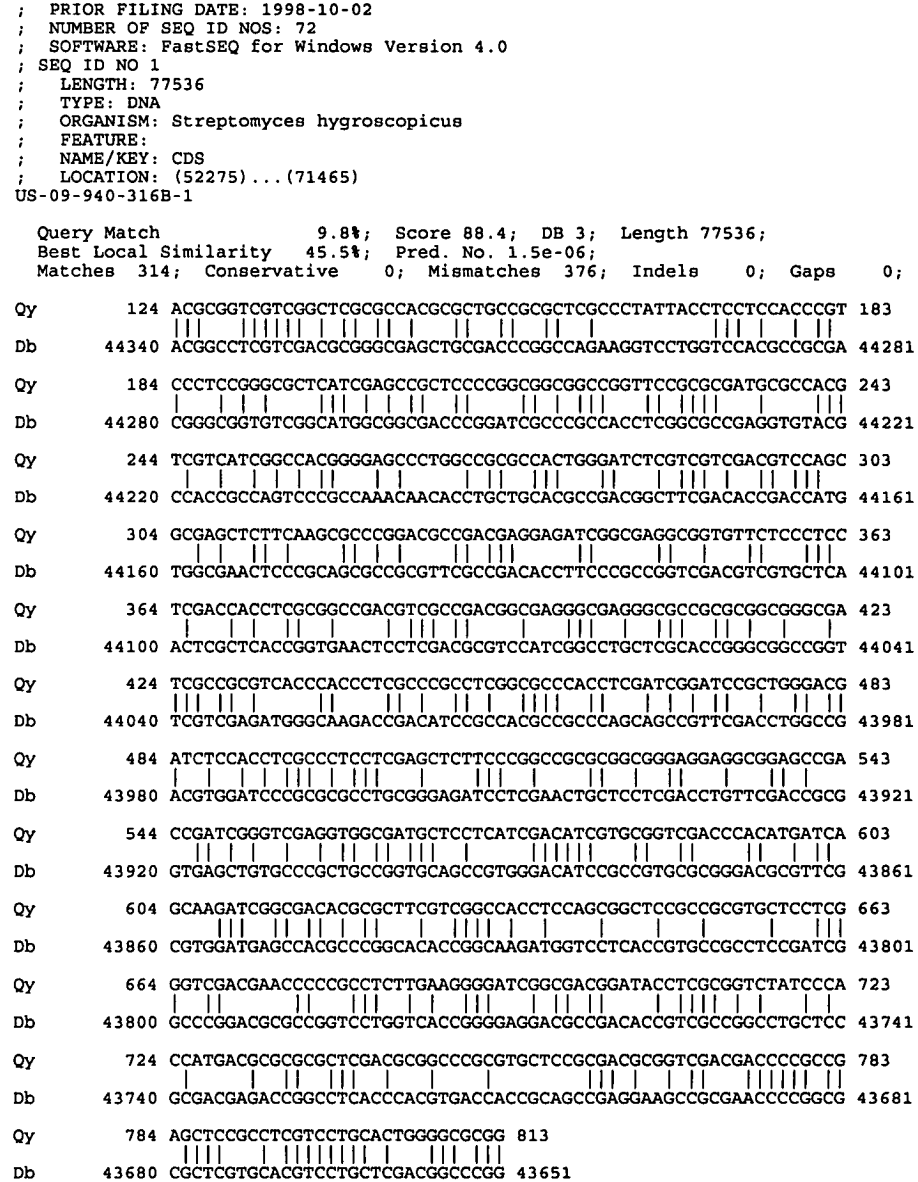

RESULT 3
US-09-410-551B-1/c
; Sequence 1, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
;  APPLICANT: KOSAN BIOSCIENCES, Inc.
;  APPLICANT:  REEVES, CHRISTOPHER
;  APPLICANT:  CHU, DANIEL
;  APPLICANT:  KHOSLA, CHAITAN
;  APPLICANT:  SANTI, DANIEL
;  APPLICANT:  WU, KAI
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE: 30062-20026.00
;  CURRENT APPLICATION NUMBER: US/09/410,551B
;  CURRENT FILING DATE:  1999-10-01
;  PRIOR APPLICATION NUMBER: US 60/139,650
;  PRIOR FILING DATE: 1999-06-17
;  PRIOR APPLICATION NUMBER: US 60/123,810
;  PRIOR FILING DATE: 1999-03-11
;  PRIOR APPLICATION NUMBER: US 60/102,748
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 77536
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (52275)...(71465)
US-09-410-551B-1

  Query Match              9.8%;  Score 88.4;  DB 3;  Length 77536;
  Best Local Similarity   45.5%;  Pred. No. 1.5e-06;
  Matches  314;  Conservative    0;  Mismatches  376;  Indels    0;  Gaps    0;
```



RESULT 4
US-09-940-316B-1/c
; Sequence 1, Application US/09940316B
; Patent No. 6759536
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
; TITLE OF INVENTION: GENE CLUSTER
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-940-316B-1

Query Match 9.8%; Score 88.4; DB 3; Length 77536;
Best Local Similarity 45.5%; Pred. No. 1.5e-06;
Matches 314; Conservative 0; Mismatches 376; Indels 0; Gaps 0;



RESULT 5
US-09-477-962-1
; Sequence 1, Application US/09477962
; Patent No. 6927286
; GENERAL INFORMATION:
; APPLICANT: SHEN, BEN
; APPLICANT: DU, LIANGCHENG
; APPLICANT: SANCHEZ, CESAR

```
;  APPLICANT:  CHEN, MEI
;  APPLICANT:  EDWARDS, DANIEL J.
;  TITLE OF INVENTION: BLEOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES
;  FILE REFERENCE: 407T-895820US
;  CURRENT APPLICATION NUMBER: US/09/477,962
;  CURRENT FILING DATE:  2000-01-05
;  PRIOR APPLICATION NUMBER: 60/115,435
;  PRIOR FILING DATE: 1999-01-06
;  PRIOR APPLICATION NUMBER: 60/118,848
;  PRIOR FILING DATE: 1999-02-05
;  NUMBER OF SEQ ID NOS: 133
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1
;   LENGTH: 58857
;   TYPE: DNA
;   ORGANISM: Streptomyces verticillus
US-09-477-962-1

  Query Match             9.5%;  Score 85.4;  DB 3;  Length 58857;
  Best Local Similarity   45.7%;  Pred. No. 4.8e-06;
  Matches  379;  Conservative    0;  Mismatches  441;  Indels    9;  Gaps    2;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              | | || | || || || |  | |   || | | | |  ||        | ||||| |||
Db      39981 CCGGCGCGGCGAGGGCGCCAACCACGCCCTGCTGCTCTCGCCGGAGCTGACCGGCCGGCT 40040

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              ||  |   |  || ||    ||  |    || |    |||     ||  |  ||  |
Db      40041 CGCCGACCTGCGCCGCCGCGAGGGCGGGTCGCTGTTCATGCTCGTGCTCTCCGCGCTCCT 40100

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              | ||   |    ||| |  ||  |  |   | || | |   |     |||| ||  |
Db      40101 GGTCGTCCTGCGTGGCACCGGCGGCCGGGACCGGCTCGCCGTCGGCACCCTCGTCGCCGG 40160

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              |  |||  |      ||||||||||  ||||  |  |      |  | | |||  |
Db      40161 CCGCACCCGCCCCGAACTCGAGCCGCTCATCGGCTACTTCGTCAACGTCCTGCTGCTGCC 40220

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              | |||  | ||||  | ||    |||   |     |  |||   |  |||  || |   |
Db      40221 CTTCGAGACCGGCGGCCGGACCTCCTTCGCCGAGCTGTGGCGGCGGGTCCGCGGCCGGCT 40280

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
              |   |||   | |   | || |||          |||||  | |  || | || |   |
Db      40281 GGTGGAGGCGTACGCCCACCAGGAACTGCCGCTGGAGAAGGCCCTGGAGCTGCTGCGCGC 40340

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              |  || |  |  | | ||||||     | | |||||  || |    | ||||  || ||
Db      40341 CGACGGCACCGCCCCCGCCGACCCGCCGGTCGGCGTGGTCTGCGTCGCCCAGCAGC---- 40396

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              ||||  |  |   ||||||  || | |||| |||   | ||||     ||  |  ||
Db      40397 --CCGCCCCCGCGATCACCCTGCCCGGACTCGACGCGAGCGTCGAGGACGTCGACCTGGG 40454

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              |    |||  ||| |||| |||   ||     |  ||  |||| ||| |  | | ||
Db      40455 CACCGCCCAGTTCGACCTCGTCG---TCGAGGTGCGCGAACGGCCGGAAGGCGTGCAGAT 40511

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              ||   | |  |||   ||  || ||  |  | || ||  ||  |||     |  |
Db      40512 CGCCTTCCAGTACGACCGGGACCTGTTCGACGCGGCCACGGTCCGGCTCCTCGCCGACCA 40571

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              |    |  || |  |   |    || || | ||| |   | || | |   || ||| |
Db      40572 CGTGCACGCCGTCCTCGACCAGGCCGCCGCCGACCCCACCCTGCCCTGTGCCGAGCTGCC 40631

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||   || ||  ||||     |  |    | |  | || ||         |  |   |
Db      40632 CGCCCCGCCGGCCCCCGCGGCCCCGGCCCGCACGGCCGGCGCCACGACGCTGCACGCCCT 40691

Qy        722 CACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGC 781
              |  | | |||||      |      ||  || | ||  | ||  | || || || | |
Db      40692 GTTCGAGTCCCGCGCCGCGAAGAGCCCCGACGCGGTCGCCCTCGTCGACGGCGGCCACCG 40751

Qy        782 CGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCAC 830
              ||    |  ||   |||  ||   ||||   ||| |   |   |||
Db      40752 CGTCACCTACCGGACCCTCAACACCCGCGCCAACCGGCTCGCCCGCCAC 40800


RESULT 6
US-07-736-178C-1
; Sequence 1, Application US/07736178C
; Patent No. 5514544
;  GENERAL INFORMATION:
;    APPLICANT:  Rao, Ramachandra N
;    APPLICANT:  Turner, Jan R
;    TITLE OF INVENTION:  ACTIVATOR GENE FOR MACROLIDE
;    TITLE OF INVENTION:  BIOSYNTHESIS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly And Company
;      STREET:  Lilly corporate Center
;      CITY:  Indianapolis
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/736,178C
;      FILING DATE:  19910726
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conrad, Robert A
;      REGISTRATION NUMBER:  32089
;      REFERENCE/DOCKET NUMBER:  X8144
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3334
;      TELEFAX:  317-276-1294
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2312 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  345..2312
US-07-736-178C-1

  Query Match             9.2%;  Score 83.2;  DB 2;  Length 2312;
  Best Local Similarity   45.2%;  Pred. No. 1.3e-05;
  Matches  381;  Conservative    0;  Mismatches  458;  Indels    3;  Gaps    2;

Qy         44 CTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCA 103
              | || |||    |||| ||  |  | |    | |   |  | |||||   |    |
Db       1154 CCTCCAGTCGCACCTGGTCACGCAGGTGATGGACAGGCGCGGCGCCGACTCGTTACTCGC 1213

Qy        104 GCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCC 163
              |      | | || ||| || ||  |  | ||||  | ||  | ||||   |||| |||
Db       1214 GACGGCCGCCGAGGCGCTCGGCGGCGGAGCCGGCCTGTGCAGCCCGCTCGGGCGCCCGCT 1273

Qy        164 CTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCG 223
              |   |   |    || |  || |   || | ||  || | | | |  || ||||
Db       1274 CGCCGAGTACGGGACCCTGCGCCCCGTCGCCCCCACGGAACTGCGCGCGGCGTGCCGCCG 1333

Qy        224 GTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGA 283
              |  |  || ||   |||    |  |  ||     |  || |  |||| ||   |  |
Db       1334 GGCCGCCGAGACCGGCCGGCCCACCTCCGTGGCCCCGGGGGTCTGGACGGTGCCCCTGCT 1393

Qy        284 TCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGG 343
              || ||  | | ||| |     |  |||   | | | |      | ||  ||  | | ||
```

```
Db       1394 TCCCGGGGGCAACGCCGGCTTCCTGCTGACCGACCTCGGTCCGGACGCGGACCACACCGC 1453

Qy        344 CGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGA 403

Db       1454 CGTCCCCCTGCTCCCGATGGTCGCCCGCACCCTCGCGCTGCACCTGCGCGTCCAGCACGA 1513

Qy        404 GGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCT 463

Db       1514 CGACTCCCCCAAGGCGCAGAGCCACCAGGAGTTCTTCGACGACCTGATCGGGGCGCCCCG 1573

Qy        464 CGATCGGATCCGCTGGGACGATCTCCACCT--CGCCCTCCTCGAGCTCTTCCCGGCCGCG 521

Db       1574 CTCACCCACGCTCCTCAGGGAACGCGCCCTGATGTTCTCCCTCAGCTTCCGCCGCCCGCA 1633

Qy        522 CGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGC-GATGCTCCTCATCGACAT 580

Db       1634 CGTGGTGCTGGTGGCGGGCGGACCCCGCGGGACCTCGCCGCGGCTGGACCGGTCCGGCGC 1693

Qy        581 CGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCA 640

Db       1694 CGACTACGCGAAGGAGCTCGGCGGGCTGTGCAGCGTGCGGGACGGCGCCGTCGTCCTGCT 1753

Qy        641 GCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGA 700

Db       1754 GCTGCCCGGCGACGACCCCGTCGCCGTGGCGCAGACCGCCGCCCCGGAGCTGACCGACCG 1813

Qy        701 CGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCG 760

Db       1814 CGCCGGGCACCCCGTCACCGTGGGGGTCGCGGGCCCCGCCTCGACCGTCGACGGCATCGC 1873

Qy        761 CGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGAT 820

Db       1874 CGACGCGCACCGTGAGGCCGCGAAGTGTCTGGAGACCCTCCGCGCGCTCGGCGGCGACGG 1933

Qy        821 GAGCGATCACGACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCAC 880

Db       1934 CGGCACCGCGTGCGCCTCCGACCTGGGTTTCCTCGGCATGCTCCTCGCCGAGGAGAACGA 1993

Qy        881 CG 882

Db       1994 CG 1995
```

RESULT 7
US-09-000-064-1
; Sequence 1, Application US/09000064
; Patent No. 6214592
; GENERAL INFORMATION:
; APPLICANT: Crouzet, Joel
; APPLICANT: Favre-Bulle, Olivier
; APPLICANT: Jourdat, Catherine
; APPLICANT: Le Coq, Anne-Marie
; APPLICANT: Petre, Dominique
; TITLE OF INVENTION: ENZYMES AND MICROORGANISMS HAVING AMIDASE ACTIVITY FOR
; TITLE OF INVENTION: HYDROLYSING POLYAMIDES
; FILE REFERENCE: 003025-036
; CURRENT APPLICATION NUMBER: US/09/000,064
; CURRENT FILING DATE: 1998-04-03
; EARLIER APPLICATION NUMBER: PCT/FR96/01119
; EARLIER FILING DATE: 1996-07-17
; EARLIER APPLICATION NUMBER: FR 9508916
; EARLIER FILING DATE: 1995-07-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1068
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1065)
US-09-000-064-1

Query Match 9.2%; Score 82.6; DB 3; Length 1068;
Best Local Similarity 46.2%; Pred. No. 1.6e-05;
Matches 274; Conservative 0; Mismatches 319; Indels 0; Gaps 0;

```
Qy        191 GGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCAT 250

Db        201 GGGGCCCACCGGTGCGACCGTGATCCACATCCCCGCCGGCGCCCGCACCGCGGTGGACGC 260

Qy        251 CGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCT 310

Db        261 GCGGGGCGGGGCGGTGGGGCTCTCCGGCGGCTACGACTTCAACCACGCCATCTGCCTGGC 320

Qy        311 CTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCA 370

Db        321 CGGCGGAGCCTGCTACGGGCTCGAGGCGGGCGCCGGGGTGAGCGACGCGCTCCTGGAACG 380

Qy        371 CCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGC 430

Db        381 CCTCGAGCATCGCACCGGCTTCGCCGAGCTCCAGCTGGTGTCGTCGGCGGTCATCTACGA 440

Qy        431 GTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCA 490

Db        441 CTTCTCGGCGCGCTCCACCCGCGTCTACCCCGACAAGGCGCTCGGCCGCGCGGCGCTCGA 500

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550

Db        501 ATTCGCCGTTCCCGGTGAGTTCCCGCAGGGGCGGGCGGGCGCGGGCATGAGCGCGTCCGC 560

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610

Db        561 GGGCAAGGTGGACTGGGACCGCACCGAGATCACCGGGCAGGGCGCGGCGTTCCGTCGTCT 620

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGAC 670

Db        621 CGGCGACGTGCGCATCCTCGCCGTCGTCGTGCCGAACCCGGTCGGTGTGATCGTGGACCG 680

Qy        671 GAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGAC 730

Db        681 CGCGGGCACGGTGGTGCGCGGCAACTACGACGCGCAGACCGGGGTCCGGCGCCACCCGGT 740

Qy        731 GCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCG 783

Db        741 GTTCGACTACCAGGAGGCGTTCGCCGAGCAGGTCCCGCCCGTCACCGAGGCCG 793
```

RESULT 8
US-10-152-886-84
; Sequence 84, Application US/10152886
; Patent No. 6912470
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING
; TITLE OF INVENTION: STRUCTURES
; FILE REFERENCE: 3011-3US
; CURRENT APPLICATION NUMBER: US/10/152,886
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84
; LENGTH: 5802
; TYPE: DNA
; ORGANISM: Streptomyces kaniharaensis
US-10-152-886-84

Query Match 9.1%; Score 81.8; DB 3; Length 5802;
Best Local Similarity 45.9%; Pred. No. 2.1e-05;
Matches 361; Conservative 0; Mismatches 417; Indels 9; Gaps 2;

```
Qy         98 GGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCG 157
```

```
Db       5019 GCTGCCGCGCGGCATCGAGCGGCTCCACCTCGCCGAACGGGCCGACCAGGACTCCGAGTT 5078

Qy        158 CTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGG 217

Db       5079 CATCGTCCTCGACGCCCGCGAACGCTCCCAGGACGGCGACACCTACGTCTACGACGTCGA 5138

Qy        218 CGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCA 277

Db       5139 CGTCCGCACCCCGGACGGCCGGGTCGTCGAACGCTGGGAGGGGCTGGCCCTGGTCGCCGT 5198

Qy        278 CTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGA 337

Db       5199 CCGCAAGCGCGACGGCGCCGGGCCGTGGGTCCCCGCGATGCTC---GGCTCCTACCTGGA 5255

Qy        338 GATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGA 397

Db       5256 ACGCGGCCTGGAG------CGCGTCCTCGGCGGCAGCCGCGCCGTCGTGGTCGAGGCCGC 5309

Qy        398 GGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGC 457

Db       5310 GGGCGCCGAAGCCGACCGGCGGGCCCGCACCGCCGACGCCGTCGCCCGGGCCCTCGGCGC 5369

Qy        458 CCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGC 517

Db       5370 ACCCGCCGAACTGCGCCACCGCCCGGACGGTCGGCCCGAACTCGACGGCCACACCGTCTC 5429

Qy        518 CGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGA 577

Db       5430 CGCCGCGCACAGCGACGGGCTGACCCTCGCCGTGGTCGGGCAGGGCCGTCTCGCCTGCGA 5489

Qy        578 CATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCT 637

Db       5490 CGCCGAGACGGTCCGCCCGCGCCGGGCCGAGGACTGGGCCGCACTCCTCGGCGAGGCCCA 5549

Qy        638 CCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGG 697

Db       5550 ACTCCCCGTCCGCGACCTGCTGGTGGCCGAGGCCGGCGACGACCCGGCGGTGGCGGCCAC 5609

Qy        698 CGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCT 757

Db       5610 CCGGGTCTGGTGCGCCCTGGAGTGCCTGCGCAAGTCCGCCGCCACCGGCCAGGCCCTCGC 5669

Qy        758 CCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCG 817

Db       5670 CCTCGACCGGGTCGACGGCTCCGGCTGGGCGGTGCTCTCCGCCGGCGACGCGGCGATCGC 5729

Qy        818 GATGAGCGATCACGACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGT 877

Db       5730 CACCTGGGTGACCACCGTCACCGACCGCGAGGACCCGGTGGTCTTCGCCTTCCTCGCCGG 5789

Qy        878 CACCGAG 884

Db       5790 CAAGGAG 5796


RESULT 9
US-09-000-040-4
; Sequence 4, Application US/09000040B
; Patent No. 6180388
; GENERAL INFORMATION:
;  APPLICANT: CROUZET, Joel
;  APPLICANT: FAUCHER, Didier
;  APPLICANT: FAVRE-BULLE, Olivier
;  APPLICANT: JOURDAT, Catherine
;  APPLICANT: PETRE, Dominique
;  APPLICANT: PIERRARD, Jerome
;  APPLICANT: THIBAUT, Denis
;  APPLICANT: GUITTON, Carole
;  TITLE OF INVENTION: ENZYMES AND MICRO-ORGANISMS WITH AMIDASE ACTIVITY WHICH
;  TITLE OF INVENTION: HYDROLYSE POLYAMIDES
;  FILE REFERENCE: 003025-035
;  CURRENT APPLICATION NUMBER: US/09/000,040B
;  CURRENT FILING DATE: 1998-04-03
;  EARLIER APPLICATION NUMBER: FR 95/08917
;  EARLIER FILING DATE: 1995-07-18
;  EARLIER APPLICATION NUMBER: PCT/FR96/01118
;  EARLIER FILING DATE: 1996-07-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1068
;   TYPE: DNA
;   ORGANISM: Comamonas acidovorans
US-09-000-040-4

  Query Match             8.8%;  Score 79.4;  DB 3;  Length 1068;
  Best Local Similarity   45.9%;  Pred. No. 5.6e-05;
  Matches  272;  Conservative    0;  Mismatches  321;  Indels    0;  Gaps    0;

Qy        191 GGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCCGATGCGCCACGTCGTCAT 250

Db        201 GGGGCCCACCGGTGCGACCGTGATCCACATCCCCGCCGGCGCCCGCACCGCGGTGGACGC 260

Qy        251 CGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCT 310

Db        261 GCGGGGCGGGGCGGTGGGGCTCTCCGGCGGCTACGACTTCAACCACGCCATCTGCCTGGC 320

Qy        311 CTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCA 370

Db        321 CGGCGGAGCCTGCTACGGGCTCGAGGCGGGCGCCGGGGTGAGCGACGCGCTCCTGGAACG 380

Qy        371 CCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGC 430

Db        381 CCTCGAGCATCGCACCGGCTTCGCCGAGCTCCAGCTGGTGTCGTCGGCGGTCATCTACGA 440

Qy        431 GTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCA 490

Db        441 CTTCTCGGCGCGCTCCACCGCGGTCTACCCCGACAAGGCGCTCGGCCGCGCGGCGCTCGA 500

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550

Db        501 ATTCGCCGTTCCCGGTGAGTTCCCGCAGGGGCGGGCGGGCGCGGGCATGAGCGCGTCCGC 560

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610

Db        561 GGGCAAGGTGGACTGGGACCGCACCGAGATCACCGGGCAGGGCGCGGCGTTCCGTCGTCT 620

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGAC 670

Db        621 CGGCGACGTGCGCATCCTCGCCGTCGTCGTGCCGAACCCGGTCGGTGTGATCGTGGACCG 680

Qy        671 GAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGAC 730

Db        681 CGCGGGCACGGTGGTGCGCGGCAACTACGACGCGCAGACCGGGGTCCGGCGCCACCCGGT 740

Qy        731 GCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCG 783

Db        741 GTTCGACTACCAGGAGGCGTTCGCCGAGCAGGTCCCGCCCGTCACCGAGGCCG 793


RESULT 10
US-09-252-991A-9713
; Sequence 9713, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
```

; SEQ ID NO 9713
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9713

Query Match 8.8%; Score 79; DB 3; Length 1239;
Best Local Similarity 46.2%; Pred. No. 6.5e-05;
Matches 351; Conservative 0; Mismatches 390; Indels 18; Gaps 2;

```
Qy         58 TGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGC 117
Db        207 TGCGCGTCGAGGGTGCCGCCGACTGGCTCGACGCGGCCCGTTTGCCGCGCGACCCGGAGG 266

Qy        118 CGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCC 177
Db        267 CAGCGCTGGCGATTCACGTGACGCCCTCGGCGCGTGCGCCGCGGGCCGAGGAACTGCTGA 326

Qy        178 ACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGC 237
Db        327 TCCATCCGCGCTGCGTGCTGGCGGCGCTCGAACCTGCCGACGCGGCGGCGGACGCGGTGC 386

Qy        238 GCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACG 297
Db        387 GCCAGGCGCTGGTCGATGCCGGCCTGGCCGCGGCGGCGCTCGGCGGCTTGCTGGCCGACA 446

Qy        298 TCCAGCGCGAGCTCTTCAAGCGCCCGGA---CGCCGACGAGGAGATCGGCGAGGCGGTGT 354
Db        447 GCGCGCGGATGGCCGATCCGCGCCTGCAGGCCGCCGCCGATGCGCTCGGCGTACCGCTGC 506

Qy        355 TCTCCCTCCTCGACCACCTCGCGGCCGA---------------CGTCGCCGACGGCGAGG 399
Db        507 GCTTCCTGCCGCAGGCCGATCCCGCTGCGCGCCTGCGCCAGGCCTTGCCCGCCGCCGAGC 566

Qy        400 GCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCC 459
Db        567 AGGTCGGCGACCGGCTGGCGGTGGCCGCCGCGCCCTTGGCGATGGACCGCATCGGCCAGC 626

Qy        460 ACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCG 519
Db        627 GTCGCGGCCGACTCACAGTGATCGGCCTGGGGCCCGGCTGCGCCGAGCTGATGGTGCCTG 686

Qy        520 CGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACA 579
Db        687 CGGCGCGGCGGGCACTGGACGAGGCCGAGGACATCCTCGGTTACCAGACCTACGTCGACA 746

Qy        580 TCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCC 639
Db        747 TGGCCGGACCCTTCCGCGCCGGACAGGTGCTGCATGGCAGCGACAACCGCGAGGAATTGC 806

Qy        640 AGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCG 699
Db        807 AGCGCGCCCGGCATGCCTTCGAACTGGCGGCGACGGGCAGGGCGGTGGTGCTGGTATCGT 866

Qy        700 ACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCC 759
Db        867 CCGGCGACCCCGGCGTGTTCGCCATGGCCGCCGCGGTGCTGGAAGCCCTGCATGGCGCTG 926

Qy        760 GCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCC 798
Db        927 GCGAGCCGGCCTGGGACGCGGTCGAGCTGCGGATCTTCC 965
```

RESULT 11
US-09-252-991A-9632
; Sequence 9632, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9632
; LENGTH: 1719
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9632

Query Match 8.8%; Score 79; DB 3; Length 1719;
Best Local Similarity 46.2%; Pred. No. 6.4e-05;
Matches 351; Conservative 0; Mismatches 390; Indels 18; Gaps 2;

```
Qy         58 TGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGC 117
Db        536 TGCGCGTCGAGGGTGCCGCCGACTGGCTCGACGCGGCCCGTTTGCCGCGCGACCCGGAGG 595

Qy        118 CGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCC 177
Db        596 CAGCGCTGGCGATTCACGTGACGCCCTCGGCGCGTGCGCCGCGGGCCGAGGAACTGCTGA 655

Qy        178 ACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGC 237
Db        656 TCCATCCGCGCTGCGTGCTGGCGGCGCTCGAACCTGCCGACGCGGCGGCGGACGCGGTGC 715

Qy        238 GCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACG 297
Db        716 GCCAGGCGCTGGTCGATGCCGGCCTGGCCGCGGCGGCGCTCGGCGGCTTGCTGGCCGACA 775

Qy        298 TCCAGCGCGAGCTCTTCAAGCGCCCGGA---CGCCGACGAGGAGATCGGCGAGGCGGTGT 354
Db        776 GCGCGCGGATGGCCGATCCGCGCCTGCAGGCCGCCGCCGATGCGCTCGGCGTACCGCTGC 835

Qy        355 TCTCCCTCCTCGACCACCTCGCGGCCGA---------------CGTCGCCGACGGCGAGG 399
Db        836 GCTTCCTGCCGCAGGCCGATCCCGCTGCGCGCCTGCGCCAGGCCTTGCCCGCCGCCGAGC 895

Qy        400 GCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCC 459
Db        896 AGGTCGGCGACCGGCTGGCGGTGGCCGCCGCGCCCTTGGCGATGGACCGCATCGGCCAGC 955

Qy        460 ACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCG 519
Db        956 GTCGCGGCCGACTCACAGTGATCGGCCTGGGGCCCGGCTGCGCCGAGCTGATGGTGCCTG 1015

Qy        520 CGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACA 579
Db       1016 CGGCGCGGCGGGCACTGGACGAGGCCGAGGACATCCTCGGTTACCAGACCTACGTCGACA 1075

Qy        580 TCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCC 639
Db       1076 TGGCCGGACCCTTCCGCGCCGGACAGGTGCTGCATGGCAGCGACAACCGCGAGGAATTGC 1135

Qy        640 AGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCG 699
Db       1136 AGCGCGCCCGGCATGCCTTCGAACTGGCGGCGACGGGCAGGGCGGTGGTGCTGGTATCGT 1195

Qy        700 ACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCC 759
Db       1196 CCGGCGACCCCGGCGTGTTCGCCATGGCCGCCGCGGTGCTGGAAGCCCTGCATGGCGCTG 1255

Qy        760 GCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCC 798
Db       1256 GCGAGCCGGCCTGGGACGCGGTCGAGCTGCGGATCTTCC 1294
```

RESULT 12
US-09-252-991A-9533/c
; Sequence 9533, Application US/09252991A
; Patent No. 6551795

```
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 9533
;   LENGTH: 3654
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9533

  Query Match             8.8%;  Score 79;  DB 3;  Length 3654;
  Best Local Similarity   46.2%;  Pred. No. 6.2e-05;
  Matches  351;  Conservative    0;  Mismatches  390;  Indels   18;  Gaps    2;

Qy         58 TGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGC 117
Db       2663 TGCGCGTCGAGGGTGCCGCCGACTGGCTCGACGCGGCCCGTTTGCCGCGCGACCCGGAGG 2604

Qy        118 CGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCC 177
Db       2603 CAGCGCTGGCGATTCACGTGACGCCCTCGGCGCGTGCGCCGCGGGCCGAGGAACTGCTGA 2544

Qy        178 ACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGC 237
Db       2543 TCCATCCGCGCTGCGTGCTGGCGGCGCTCGAACCTGCCGACGCGGCGGCGGACGCGGTGC 2484

Qy        238 GCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACG 297
Db       2483 GCCAGGCGCTGGTCGATGCCGGCCTGGCCGCGGCGGCGCTCGGCGGCTTGCTGGCCGACA 2424

Qy        298 TCCAGCGCGAGCTCTTCAAGCGCCCGGA---CGCCGACGAGGAGATCGGCGAGGCGGTGT 354
Db       2423 GCGCGCGGATGGCCGATCCGCGCCTGCAGGCCGCCGCCGATGCGCTCGGCGTACCGCTGC 2364

Qy        355 TCTCCCTCCTCGACCACCTCGCGGCCGA--------------CGTCGCCGACGGCGAGG 399
Db       2363 GCTTCCTGCCGCAGGCCGATCCCGCTGCGCGCCTGCGCCAGGCCTTGCCCGCCGCCGAGC 2304

Qy        400 GCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCC 459
Db       2303 AGGTCGGCGACCGGCTGGCGGTGGCCGCCGCGCCCTTGGCGATGGACCGCATCGGCCAGC 2244

Qy        460 ACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCG 519
Db       2243 GTCGCGGCCGACTCACAGTGATCGGCCTGGGGCCCGGCTGCGCCGAGCTGATGGTGCCTG 2184

Qy        520 CGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACA 579
Db       2183 CGGCGCGGCGGGCACTGGACGAGGCCGAGGACATCCTCGGTTACCAGACCTACGTCGACA 2124

Qy        580 TCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCC 639
Db       2123 TGGCCGGACCCTTCCGCGCCGGACAGGTGCTGCATGGCAGCGACAACCGCGAGGAATTGC 2064

Qy        640 AGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCG 699
Db       2063 AGCGCGCCCGGCATGCCTTCGAACTGGCGGCGACGGGCAGGGCGGTGGTGCTGGTATCGT 2004

Qy        700 ACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCC 759
Db       2003 CCGGCGACCCCGGCGTGTTCGCCATGGCCGCCGCGGTGCTGGAAGCCCTGCATGGCGCTG 1944

Qy        760 GCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCC 798
Db       1943 GCGAGCCGGCCTGGGACGCGGTCGAGCTGCGGATCTTCC 1905

RESULT 13
US-10-152-886-54
; Sequence 54, Application US/10152886
; Patent No. 6912470
; GENERAL INFORMATION:
;  APPLICANT: ECOPIA BIOSCIENCES INC.
;  APPLICANT:  Farnet, Chris
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING
;  TITLE OF INVENTION:  STRUCTURES
;  FILE REFERENCE: 3011-3US
;  CURRENT APPLICATION NUMBER: US/10/152,886
;  CURRENT FILING DATE:  2002-05-21
;  NUMBER OF SEQ ID NOS: 102
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 54
;   LENGTH: 5877
;   TYPE: DNA
;   ORGANISM: Kitasatosporia sp.
US-10-152-886-54

  Query Match             8.7%;  Score 78.8;  DB 3;  Length 5877;
  Best Local Similarity   46.0%;  Pred. No. 6.6e-05;
  Matches  352;  Conservative    0;  Mismatches  402;  Indels   12;  Gaps    2;

Qy         74 CGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGT 133
Db       3057 CAGCCAACTCGCCGACCTGCTGGACGAGTTGGCCCAGCAGTCACCGCAGGACACCCGCCC 3116

Qy        134 CGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGC 193
Db       3117 CGGCGCCGCCGCCGGCGTCGCCCCCTGGGTGCGGCCCTTCCGGATCGACCTCACCGAGAC 3176

Qy        194 GCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGG 253
Db       3177 CCCGCCGCCCGCCCCCGCCGCGCCGGACCCGGCGGCCGCTGGGAGGTCTTCGCCACCGACCG 3236

Qy        254 CCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTT 313
Db       3237 GCACCCGCTGGCCGGACCGCTCGCCGAGCGGCTCACCGCCACCGCCCCCGGCGGCGGCGT 3296

Qy        314 CAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCT 373
Db       3297 CCTGCTCGCCCTGCCCCGCGA------CTGCGACCAGCGCCACCTCGGCCTGATGCTCGC 3350

Qy        374 CGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTC 433
Db       3351 CGCCGCCCGCGCCGCCCTCGACCCGGCCCGCCGCGCGGCCGGCACCCGGCTCGTCGCCGT 3410

Qy        434 ACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCT 493
Db       3411 CGGCGACCACCGCGGCGCCGCCGGCCTCGCCAAGACGCTCCACCTGGAGGCCCCCGACAT 3470

Qy        494 CGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGT 553
Db       3471 CCCGGTCACCGTCGTCACCCTGCCCCTCGACCAGGAACTGCCCGCCCCGGCCGCCGAGCA 3530

Qy        554 CGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGG 613
Db       3531 GGCCGCCGCCCGGATCGCCGCCGACACCGCCGCCACCACCGGCTTCAGCGAGGTCCACTA 3590

Qy        614 CGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAA 673
Db       3591 CGACGCCGACGGCACCCGCCGGGTCCCCGTGCTGCGCCCCGTCCCGCTCGAACC------ 3644

Qy        674 CCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCG 733
Db       3645 CGACCCCGGACGGCAGGCCCTCGGCCCGCGGGACGTCCTGCTGGTCACCGGCGGCGGCAA 3704

Qy        734 CGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCT 793
```

```
Db       3705 GGGCATCACCGCCGAGTGCGCCCTCGCCCTGGCCGGCGGCAACGGCGCCGCGATCGGCCT 3764

Qy        794 CGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATC 839
Db       3765 GATCGGCCGCTCCGACCCCGCCCGGGACACCGAACTCGCCGACAAC 3810


RESULT 14
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             8.7%;  Score 78.2;  DB 2;  Length 44377;
  Best Local Similarity   44.9%;  Pred. No. 7.7e-05;
  Matches  385;  Conservative    0;  Mismatches  463;  Indels    9;  Gaps    2;

Qy          8 GGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGT 67
Db      22005 GGACTGGAACCTCGCGGACGTGCTGCGCGGCGCGCCCGCGGCGCCCGGCCTCGACCGGGT 22064

Qy         68 GATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGC 127
Db      22065 CGACGTCGTCCAGCCGGCCACCTTCGCCGTCATGGTGGGACTCGCCGCGCTGTGGCGCTC 22124

Qy        128 GGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCT 187
Db      22125 CCTCGGGGTCGAACCCGCCGCCGTCATCGGCCACTCCCAGGGCGAGATCGCCGCGGCCTG 22184

Qy        188 CCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGT 247
Db      22185 CGTGGCGGGCGCGCTCTCCCTGGAGGACGCCGCCCGGATCGTGGCCCTGCGCTCCCAGGT 22244

Qy        248 CATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGA 307
Db      22245 CATCGCCCGCGAACTGGCCGGGCGGGGCGGCATGGCCTCGGTGGCCCTGCCCGCGGCGGA 22304

Qy        308 GCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGA 367
Db      22305 GGTCGAGGCCCGCCTGGCCGGCGGCGTCGAGATCGCCGCCGTCAACGGCCCCGGCTCGAC 22364

Qy        368 CCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGG------GC 421
Db      22365 CGTCGTCTGCGGAGAGCCCGGCGCCCTGGAGGCGTTGCTCGTCACGCTGGAGAGCGAAGG 22424

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
Db      22425 CACCCGGGTCCGCCGCATCGACGTCGACTACGCGTCCCACTCCCACTACGTCGAGAGCAT 22484

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
Db      22485 CCGGGCGGAACTCGCCACCGTCCTCGGCCCCGTCCGGCCGCGGAGGGGCGACGTGCCCTT 22544

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
Db      22545 CTACTCCACCGTCGAGGCGGCGCTCCTCGACACCGCCACCCTGGACGCCGACTACTGGTA 22604

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
Db      22605 CCGCAACCTGCGCCTCCCGGTGCGCTTCGAGCCGACCGTACGCGCCATGCTCGACGACGG 22664

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
Db      22665 CGTCGACGCGTTCGTGGAGTGCT---CCGCGCATCCCGTCCTGACCGTCGGCGTGCGCCA 22721

Qy        722 CACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGC 781
Db      22722 GACCGTGGAGAGCGCCGGCGGCGCGGTCCCGGCCCTCGCTTCGCTGCGCCGCGACGAGGG 22781

Qy        782 CGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGG 841
Db      22782 CGGGCTGCGGCGCTTCCTCACCTCCGCCGCCGAGGCCCAGGTCGTCGGCGTCCCCGTGGA 22841

Qy        842 CAGGGAGGTCCACCGGC 858
Db      22842 CTGGGCGACGCTCCGCC 22858


RESULT 15
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
```

```
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             8.7%;  Score 78.2;  DB 2;  Length 44377;
  Best Local Similarity   44.9%;  Pred. No. 7.7e-05;
  Matches  385;  Conservative    0;  Mismatches  463;  Indels    9;  Gaps    2;

Qy          8 GGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGT 67
              ||||  | ||  ||    |     | |  | ||| || |   | |  ||||||   | ||||
Db      22005 GGACTGGAACCTCGCGGACGTGCTGCGCGGCGCGCCCGCGGCGCCCGGCCTCGACCGGGT 22064

Qy         68 GATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGC 127
                || || | | |||||  ||           ||  | |  ||  |  || |     | |
Db      22065 CGACGTCGTCCAGCCGGCCACCTTCGCCGTCATGGTGGGACTCGCCGCGCTGTGGCGCTC 22124

Qy        128 GGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCT 187
                |||  | |   | |  |||  |     ||    || |   |    ||  | || ||
Db      22125 CCTCGGGGTCGAACCCGCCGCCGTCATCGGCCACTCCCAGGGCGAGATCGCCGCGGCCTG 22184

Qy        188 CCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGT 247
              |  ||||  |       | ||    | || || |||| ||   ||  |||||  |   ||
Db      22185 CGTGGCGGGCGCGCTCTCCCTGGAGGACGCCGCCCGGATCGTGGCCCTGCGCTCCCAGGT 22244

Qy        248 CATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGA 307
              ||||| || ||    | || ||| | ||  |  ||     || | |     |  |   ||
Db      22245 CATCGCCCGCGAACTGGCCGGGCGGGGCGGCATGGCCTCGGTGGCCCTGCCCGCGGCGGA 22304

Qy        308 GCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGA 367
              | ||      |||| || || || ||  ||||||| ||  |       | ||  |
Db      22305 GGTCGAGGCCCGCCTGGCCGGCGGCGTCGAGATCGCCGCCGTCAACGGCCCCGGCTCGAC 22364

Qy        368 CCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGG------GC 421
              |  | ||   |  ||   || || |   ||||    |  || | ||| | |       |
Db      22365 CGTCGTCTGCGGAGAGCCCGGCGCCCTGGAGGCGTTGCTCGTCACGCTGGAGAGCGAAGG 22424

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              | |  |      | || |  || | |  ||   |||||  | |     ||    |
Db      22425 CACCCGGGTCCGCCGCATCGACGTCGACTACGCGTCCCACTCCCACTACGTCGAGAGCAT 22484

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              |      | ||||||  | ||      |  |        |||||  |||  || | |
Db      22485 CCGGGCGGAACTCGCCACCGTCCTCGGCCCCGTCCGGCCGCGGAGGGGCGACGTGCCCTT 22544

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
                  |  |||||||  |||| | |||  || || || | ||    |   | || |
Db      22545 CTACTCCACCGTCGAGGCGGCGCTCCTCGACACCGCCACCCTGGACGCCGACTACTGGTA 22604

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||||  |  ||  | ||     | |||  |    |   ||   |  | ||    |
Db      22605 CCGCAACCTGCGCCTCCCGGTGCGCTTCGAGCCGACCGTACGCGCCATGCTCGACGACGG 22664

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||      ||  |         ||      | |      || |      | |||  ||    |
Db      22665 CGTCGACGCGTTCGTGGAGTGCT---CCGCGCATCCCGTCCTGACCGTCGGCGTGCGCCA 22721

Qy        722 CACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGC 781
              |||  ||  | ||||   || ||||| ||  |  |||        | |  ||||   |
Db      22722 GACCGTGGAGAGCGCCGGCGGCGCGGTCCCGGCCCTCGCTTCGCTGCGCCGCGACGAGGG 22781

Qy        782 CGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGG 841
              || ||| || | | ||||   ||  | ||| | |      |  ||   || |   | |
Db      22782 CGGGCTGCGGCGCTTCCTCACCTCCGCCGCCGAGGCCCAGGTCGTCGGCGTCCCCGTGGA 22841

Qy        842 CAGGGAGGTCCACCGGC 858
              | ||| |   | ||| |
Db      22842 CTGGGCGACGCTCCGCC 22858


Search completed: April 27, 2006, 12:42:39
Job time : 186.89 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 903.317 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_75000_75900
Perfect score: 901
Sequence: 1 acggccgggacgcggacgac.....gaggaggatcgcgccgcgga 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	901	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
	2	91.4	10.1	1686	6	US-10-156-761-3852	Sequence 3852, Ap
c	3	91.4	10.1	9025608	6	US-10-156-761-1	Sequence 1, Appli
	4	90.8	10.1	1677	6	US-10-156-761-1942	Sequence 1942, Ap
	5	90.2	10.0	1433	7	US-10-425-114-15862	Sequence 15862, A
	6	90.2	10.0	1456	7	US-10-425-114-15594	Sequence 15594, A
	7	88.8	9.9	11058	6	US-10-156-761-3629	Sequence 3629, Ap
	8	88.6	9.8	1583	8	US-10-425-115-156650	Sequence 156650,
c	9	88.4	9.8	77536	3	US-09-940-316B-1	Sequence 1, Appli
	10	87.8	9.7	22533	9	US-10-760-493-36	Sequence 36, Appl
	11	87.8	9.7	164051	9	US-10-760-493-18	Sequence 18, Appl
	12	87.4	9.7	3948	7	US-10-282-122A-25736	Sequence 25736, A
	13	87	9.7	1254	7	US-10-425-114-15723	Sequence 15723, A
	14	87	9.7	1434	7	US-10-425-114-15784	Sequence 15784, A
	15	86.8	9.6	27541	7	US-10-203-295-2	Sequence 2, Appli
	16	86.8	9.6	125401	7	US-10-203-295-35	Sequence 35, Appl
c	17	86.2	9.6	1633	7	US-10-437-963-78818	Sequence 78818, A
	18	86.2	9.6	31263	7	US-10-282-122A-25447	Sequence 25447, A
	19	86	9.5	1617	6	US-10-156-761-7201	Sequence 7201, Ap
	20	86	9.5	9025608	6	US-10-156-761-1	Sequence 1, Appli
	21	84	9.3	1047	6	US-10-107-431-94	Sequence 94, Appl
	22	84	9.3	1417	8	US-10-425-115-143892	Sequence 143892,
	23	84	9.3	45055	6	US-10-107-431-277	Sequence 277, App
	24	83.8	9.3	65140	7	US-10-203-295-1	Sequence 1, Appli
	25	83.6	9.3	1323	6	US-10-156-761-6612	Sequence 6612, Ap
c	26	83.4	9.3	1337	7	US-10-437-963-57680	Sequence 57680, A
	27	83.4	9.3	1839	7	US-10-425-114-13547	Sequence 13547, A
	28	83	9.2	3771	6	US-10-156-761-5283	Sequence 5283, Ap
c	29	82.8	9.2	138203	9	US-10-819-386A-1	Sequence 1, Appli
	30	82.2	9.1	1380	7	US-10-417-700A-138	Sequence 138, App
	31	81.8	9.1	2028	8	US-10-425-115-66559	Sequence 66559, A
	32	81.8	9.1	5802	5	US-10-152-886-84	Sequence 84, Appl
	33	81.8	9.1	5802	10	US-11-053-576-84	Sequence 84, Appl
	34	81.8	9.1	5802	10	US-11-053-052-84	Sequence 84, Appl
	35	81.6	9.1	1796	7	US-10-437-963-43317	Sequence 43317, A
	36	81.6	9.1	7788	6	US-10-329-079-8	Sequence 8, Appli
	37	81.6	9.1	37360	6	US-10-329-079-6	Sequence 6, Appli
	38	81.2	9.0	3471	6	US-10-156-761-2726	Sequence 2726, Ap
	39	81	9.0	936	7	US-10-354-437-3	Sequence 3, Appli
	40	80.6	8.9	1098	7	US-10-437-963-72023	Sequence 72023, A
	41	80.6	8.9	1266	7	US-10-282-122A-15034	Sequence 15034, A
	42	80.6	8.9	1446	6	US-10-156-761-3406	Sequence 3406, Ap
	43	80.6	8.9	11817	6	US-10-156-761-2884	Sequence 2884, Ap
	44	80.6	8.9	125746	6	US-10-156-761-15102	Sequence 15102, A
c	45	80	8.9	1075	8	US-10-425-115-5563	Sequence 5563, Ap

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 901; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 8e-204;
Matches 901; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ACGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75000 ACGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGC 75059

Qy         61 TCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75060 TCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGC 75119

Qy        121 ACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75120 ACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACC 75179
```

```
Qy        181 CGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75180 CGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCC 75239

Qy        241 ACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75240 ACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCC 75299

Qy        301 AGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75300 AGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCC 75359

Qy        361 TCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75360 TCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGG 75419

Qy        421 CGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75420 CGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGG 75479

Qy        481 ACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75480 ACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGC 75539

Qy        541 CGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75540 CGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGA 75599

Qy        601 TCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75600 TCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCC 75659

Qy        661 TCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75660 TCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATC 75719

Qy        721 CCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75720 CCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCG 75779

Qy        781 CCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75780 CCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCG 75839

Qy        841 GCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75840 GCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCGCGG 75899

Qy        901 A 901
              |
Db      75900 A 75900
```

RESULT 2
US-10-156-761-3852
; Sequence 3852, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3852
; LENGTH: 1686
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1686)
US-10-156-761-3852

Query Match 10.1%; Score 91.4; DB 6; Length 1686;
Best Local Similarity 47.4%; Pred. No. 2.3e-12;
Matches 420; Conservative 0; Mismatches 446; Indels 21; Gaps 4;

```
Qy         15 GACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGC 74
              | ||||   |  || |  ||| |||| |  ||    |||||   || |  ||  |   |
Db         52 GGCGACCGTTTCGACGCCCTCCACGAGATGCTCACCGGCGGCGAACTGGACGTCCTGACC 111

Qy         75 GGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCG------CCAGCCGCACGACGCG 128
              |||||       ||||||| ||   |||    | || |       ||  | | | ||| |
Db        112 GGCGACTACCTCGCCGAGCTGACCATGCTCATCCTCGGCCGCGACCGCCTGAAGGACCCC 171

Qy        129 GTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTC 188
              || ||| |    || | |  | |  ||| || | |   |  |||| ||   ||| |
Db        172 ACCGCCGGTTACGCCCGCACCTTCCTCCGCCAGCTCGAGGGCTGCCTCGGCCTCGCCCAC 231

Qy        189 CGGGCGCTCATCGAGCCGCTCCC---CGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              | | | | |    || |  ||  |||||||    || |||    | |  || |
Db        232 GAGCGGGGCGTGCGGATCGTCACCAACGCGGGCGGCCTCAACCCCGCCGGACTCGCCGAC 291

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
              | |||| |  ||  ||       | |   ||  |       | ||||| | ||||  || ||
Db        292 GCCATCCGGAACCTGGCCGGACGCCTCGGCATCCCCACCCGCGTCGCCCACGTGGAGGGC 351

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
              ||   | |||    | || || | | |         || | | ||     ||  |  | ||
Db        352 GACGACCTCACCGCCGCGCACCCGGGCAGCCTCGCCGCCCACGCCTACCTCGGCGGCTTC 411

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              | |  | ||||       ||  | || |||   || || | |||   ||||| || |
Db        412 GGCATCGCCGCGTGTCTGCGCGAGGGCGCCGACATCGTGGTCACCG---GGCGGGTGACC 468

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
              | |||| | |    | || | |  || ||||||| |||   ||    |   |||| |
Db        469 GACGCGGCCCTGGTCACCGGGCCCGCCGCCGCCCACTTCGGCTGGCAGCCGGGGGAGTAC 528

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              ||  || ||   |  ||   ||  |  |  || || |   |   | |||| || | |||
Db        529 GACCGGCTGGCGGGCGCCGTCGTCGCCGGGCACGTGCTGGAGTGCGGGGCGCAGGCCACC 588

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
              |   ||  | | |      | | |  |   ||| ||   ||  |     |  | ||
Db        589 GGCGGGAACTACGCCTTCTTCCGCGACGCGGACGTCCGCCGCCCCGGGTTCCCGCTCGCG 648

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGC---------TCCGCCGCGTG 656
              |||||  ||       ||  || || |  |  | |||  |           | ||| |||
Db        649 GAGATCCACGCGGACGGCACCGGCGTCATCACCAAGCACCCCGGCACCGGCGGCCTCGTC 708

Qy        657 CTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTC 716
              | ||||  ||   | | |||   || || | | ||  |  | | ||        |||| |
Db        709 GACGTCGGTACGGTCACCGCCCAGTTCCTGTACGAGACGGCGGGCGCCCGGTACGCGGGC 768

Qy        717 TATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGAC 776
                |  ||   | |||     |  ||      | |   |   || | | | |  || |
Db        769 CCCGACGTCACCGCCCGCCTCGACACCGTACGGCTCACCCAGGACGGCCCCGACCGCGTC 828

Qy        777 CCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTG 836
              | || ||||  |   | || | |  |  | | ||| |  | |  ||  | | ||  |
Db        829 CGCGTCGAGGGCGTACGCGGCGAGGCCCCGCCCCCGGCCCTCAAGGTCGGCCGCAACCGG 888

Qy        837 ATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGA 883
              |||||  |      | ||  | || ||    ||  | | |    |||
```

```
Db        889 CTCGGCGGCTTCCGCAACGAGGTCGTCTTCGTCCTGACCGGTCTCGA 935


RESULT 3
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             10.1%;  Score 91.4;  DB 6;  Length 9025608;
  Best Local Similarity   47.4%;  Pred. No. 6.3e-13;
  Matches  420;  Conservative    0;  Mismatches  446;  Indels   21;  Gaps    4;

Qy         15 GACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGC 74
              | ||||   |   || |  ||| |||| |  ||    |||||   || |  ||  |   |
Db    4779894 GGCGACCGTTTCGACGCCCTCCACGAGATGCTCACCGGCGGCGAACTGGACGTCCTGACC 4779835

Qy         75 GGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCG------CCAGCCGCACGACGCG 128
              |||||       ||||||| ||   |||    | || |      ||  | | | ||| |
Db    4779834 GGCGACTACCTCGCCGAGCTGACCATGCTCATCCTCGGCCGCGACCGCCTGAAGGACCCC 4779775

Qy        129 GTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTC 188
              || ||| |    || | |  | |  ||| || | |   |  |||| ||   ||| |
Db    4779774 ACCGCCGGTTACGCCCGCACCTTCCTCCGCCAGCTCGAGGGCTGCCTCGGCCTCGCCCAC 4779715

Qy        189 CGGGCGCTCATCGAGCCGCTCCC---CGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              |  | | |  |   || |   ||  |||||||    || |||    | |  || |
Db    4779714 GAGCGGGGCGTGCGGATCGTCACCAACGCGGGCGGCCTCAACCCCGCCGGACTCGCCGAC 4779655

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
              | |||| |  ||  ||      | |   ||  |       | ||||| | ||||  || ||
Db    4779654 GCCATCCGGAACCTGGCCGGACGCCTCGGCATCCCCACCCGCGTCGCCCACGTGGAGGGC 4779595

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
              ||   | |||    | || || | | |          || | | ||      ||  |  | ||
Db    4779594 GACGACCTCACCGCCGCGCACCCGGGCAGCCTCGCCGCCCACGCCTACCTCGGCGGCTTC 4779535

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              | |  |  ||||       ||  | || |||   || || | |||   ||||| || | |
Db    4779534 GGCATCGCCGCGTGTCTGCGCGAGGGCGCCGACATCGTGGTCACCG---GGCGGGTGACC 4779478

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
              | |||| | |     |  ||  | |  || ||||||| |||   ||    |   ||||  |
Db    4779477 GACGCGGCCCTGGTCACCGGGCCCGCCGCCGCCCACTTCGGCTGGCAGCCGGGGGAGTAC 4779418

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              ||  || ||   |  ||   ||  |  |  || || |  |   | |||| || | |||
Db    4779417 GACCGGCTGGCGGGCGCCGTCGTCGCCGGGCACGTGCTGGAGTGCGGGGCGCAGGCCACC 4779358
```

```
Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
              |   ||  | | |      | | |  |  ||| ||   ||  |      |  | ||
Db    4779357 GGCGGGAACTACGCCTTCTTCCGCGACGCGGACGTCCGCCGCCCCGGGTTCCCGCTCGCG 4779298

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGC---------TCCGCCGCGTG 656
              |||||  ||       ||  || || |  |  ||| |           | ||| |||
Db    4779297 GAGATCCACGCGGACGGCACCGGCGTCATCACCAAGCACCCCGGCACCGGCGGCCTCGTC 4779238

Qy        657 CTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTC 716
              | ||||  ||  | | |||  || || | | ||  |  | ||          ||||| |
Db    4779237 GACGTCGGTACGGTCACCGCCCAGTTCCTGTACGAGACGGCGGGCGCCCGGTACGCGGGC 4779178

Qy        717 TATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGAC 776
                | ||   | |||      |  ||     | |   |    || | | | |  || |
Db    4779177 CCCGACGTCACCGCCCGCCTCGACACCGTACGGCTCACCCAGGACGGCCCCGACCGCGTC 4779118

Qy        777 CCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTG 836
              | || |||| |   | || |  |  |  | | |||  |  | |  ||  | | | ||  |
Db    4779117 CGCGTCGAGGGCGTACGCGGCGAGGCCCCGCCCCCGGCCCTCAAGGTCGGCCGCAACCGG 4779058

Qy        837 ATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGA 883
              ||||| |     | ||  | || ||   ||  | | |   |||
Db    4779057 CTCGGCGGCTTCCGCAACGAGGTCGTCTTCGTCCTGACCGGTCTCGA 4779011


RESULT 4
US-10-156-761-1942
; Sequence 1942, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1942
;   LENGTH: 1677
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1677)
US-10-156-761-1942

  Query Match             10.1%;  Score 90.8;  DB 6;  Length 1677;
  Best Local Similarity   45.2%;  Pred. No. 3.1e-12;
  Matches  332;  Conservative    0;  Mismatches  402;  Indels    0;  Gaps    0;

Qy         86 GGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCA 145
              | || ||| |   |  |||||  || || |    || |   ||| |||   |  | | |
Db        942 GCCCAAGCAGCGCGGAAAGCGCCGCCCCCGGGGCAGCGCCCGCTTCGCCGGGATGATCGA 1001

Qy        146 CGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCC 205
              ||||| |  |    |||||      |  ||| | | |  | | |    |      ||  ||
Db       1002 CGCGCAGGAGACACCGCCCGCCGCCACCCTGCCCGCCGCACCCGCTCCCGCGCCCGCACC 1061

Qy        206 GCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCC 265
              |  | ||||| ||  ||    | ||||      |  ||  | |  ||| || |    | |
Db       1062 GAACACCGGCCGCACCCCCCGCGGCGCCCGCTTCGCCGGAGCCGCCGGGCAGGCCCCGGC 1121

Qy        266 CTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGA 325
```

```
Db       1122 GGCGAAGCCCCGACGCGAGCCGCTGGACGACGAGGCGCGCCGCGACACCATCGAGACCGT 1181

Qy        326 CGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGT 385

Db       1182 GCGCACACTGACCCGACTGCGCGGCGAGGGCCGCAGCGGTGAGGCGCACGCGCTGCTCGT 1241

Qy        386 CGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGC 445

Db       1242 CGAGGCCGCGTACTGGCCCGCCGCCCGCTTCCCGCTGCTCGCCGCCGAACTGCACCGCGT 1301

Qy        446 CCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGA 505

Db       1302 CGGCCTCGGCGCCGACTGGGCCACCCTCCTGTGGGAGGCGGCCTCGCTCCCCGCCGACCG 1361

Qy        506 GCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGAT 565

Db       1362 CCTGGTCGCCGCCGCGGACGCGCTCGTCGCGGCGGGCCGCACGGTGGACGGGCAACAGAT 1421

Qy        566 GCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTT 625

Db       1422 GCTGCGGCAGGGGGTGGTCCGGCCCGCCCCGGAGATCGGCGAGGCCCTGCTCGGGCTCAC 1481

Qy        626 CGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTT 685

Db       1482 CGGCGACGGCCGCCACCGCGAGGCCCGCGCCCTGCTCGACGCGTACGTACGCATCCGCAC 1541

Qy        686 GAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGC 745

Db       1542 TCCGGAGGAAGCGGCCCGCAGTGTCGAGGCGGACCCGCACACGCTCGTCCCCCTGCTCCT 1601

Qy        746 GGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTG 805

Db       1602 GGAGGCGGCCCGGGGCGTCTCCGAGGAATGCCACTGGGACCTCGTCCACGCCCTCCGTGT 1661

Qy        806 GGGCGCGGTGCGGA 819

Db       1662 CGCCGGGTTCACGA 1675
```

RESULT 5
US-10-425-114-15862
; Sequence 15862, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15862
; LENGTH: 1433
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-094-H10_FLI
US-10-425-114-15862

Query Match 10.0%; Score 90.2; DB 7; Length 1433;
Best Local Similarity 46.8%; Pred. No. 4.5e-12;
Matches 407; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```
Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65

Db        275 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 334

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125

Db        335 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 394

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185

Db        395 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 451

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245

Db        452 CTGCTGAGCCAGCTCGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 511

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305

Db        512 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 571

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365

Db        572 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC----------C 619

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425

Db        620 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 679

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485

Db        680 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 730

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545

Db        731 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 790

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605

Db        791 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 850

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665

Db        851 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 910

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725

Db        911 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCCCGCAAGCCCCTC 970

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785

Db        971 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCGCACGCGCACGGCGCCGTG 1030

Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845

Db       1031 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1090

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874

Db       1091 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1119
```

RESULT 6
US-10-425-114-15594
; Sequence 15594, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28

```
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15594
;   LENGTH: 1456
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3060-034-E6_FLI
US-10-425-114-15594

  Query Match            10.0%;  Score 90.2;  DB 7;  Length 1456;
  Best Local Similarity   46.8%;  Pred. No. 4.5e-12;
  Matches  407;  Conservative    0;  Mismatches  438;  Indels   24;  Gaps    3;

Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
              ||    | | |  |||||  | | || |  ||  |||||           ||  |  |
Db        284 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 343

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
                |   | | |||     || |  | || |  || |  || || |  | | | ||  | |
Db        344 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 403

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
              ||  ||| |  | |||   ||   ||||||| ||  ||||| | |  |  | |||    |
Db        404 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 460

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              || | |   |  ||||||    ||      |   |||   |||  |    || || |  |
Db        461 CTGCTGAGCCAGCTCGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 520

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
                | ||| |  ||  |   || | |||| ||          | ||||   || |  ||||
Db        521 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 580

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
                | |||  |     | |   || | || | || ||| ||| ||  |              |
Db        581 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC------------C 628

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              || ||| ||||  |  | |||     || ||| | || |  |||||||||   |||| ||
Db        629 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 688

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
                  | ||||  || |  |  | || |  ||  || | |  |  ||  ||||
Db        689 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 739

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              ||  ||| ||     |||| |  | |  | |||| |  |  || |  |  |  |  | |
Db        740 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 799

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
                ||    |  |||||    |  |      | ||||||    | |||          ||
Db        800 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 859

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
                ||    || |     | | |  | |       || | | || || |   | |
Db        860 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 919

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
                |  | ||| ||       | ||| |||  |  ||    | ||||  | ||  |
Db        920 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCCCGCAAGCCCCTC 979

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
                |   ||     |||| | |  || | |  |  |||||  ||   ||  ||||| |
Db        980 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCGCACGCGCACGGCGCCGTG 1039

Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845
              |||  |||     |  | |||  |||  | ||  |     | | ||||  |   |||  |
Db       1040 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1099

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874
              | ||| |     ||   | | |  | ||||
Db       1100 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1128


RESULT 7
US-10-156-761-3629
; Sequence 3629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3629
;   LENGTH: 11058
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(11058)
US-10-156-761-3629

  Query Match             9.9%;  Score 88.8;  DB 6;  Length 11058;
  Best Local Similarity   47.3%;  Pred. No. 7.1e-12;
  Matches  375;  Conservative    0;  Mismatches  402;  Indels   15;  Gaps    3;

Qy         35 CCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCG 94
              |||| |   | | ||  ||| |  || ||    ||||  || ||  |||    ||  ||
Db        300 CCGCTACGCCCTGGACGGCGACGAGCCCGTCCAGATCATCGACGCGCCGCGCCCCCTGCC 359

Qy         95 GAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCG---CT 151
                ||  | |  |       |  |  ||||||||||  || |  |  | | |||   ||
Db        360 CCTGGACCGGTCCGGGGCCGCGGACGACGACGCGGTCCGCGCCCTGGTCGAGGCGGACCT 419

Qy        152 GCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCC 211
              | | ||| ||  | |   |  || | ||  | |  | |||||| | |    | |||| |
Db        420 GGCCCGCCCGTTCGACCTCGGCCGCGACTGGCCGGTGCGGGCGCGCCTGATCCGGCTCGC 479

Qy        212 CGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCC 271
              || || ||  |  ||| |||| ||  |  |  |  ||||| |  ||  |    |  |
Db        480 CGACGACGAGCACGTCCTCGCGGTGGTCTTCCACCACATCGCCTGCGACGCCTGGTCCAC 539

Qy        272 GCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGA 331
                ||  ||  |  |  |   ||  || || |  |  |  ||      |  |      |
Db        540 CGGCGTCTTCGGACAGGAACTCAGCGCCCTGTACACGGGCTCGGCCCCGGCCCCGCTCAC 599

Qy        332 CGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGA 391
              || | ||  ||  ||    |   ||  |  | || | | ||  | | ||| |||| |||
Db        600 CGTGCAGTACGCGGACTACGCCGCCTGGCAGCGCGCCGAGCTGACCGGCGAGGTCGTCGA 659

Qy        392 CGG---CGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCG 448
                  |   |  | | |    |  |||| ||||  |  ||||  | | |   ||  ||
Db        660 ACGTCACCTCGACCACTGGAAGGGCCGGCTGGCGGACCTCGCGCCGCTCGAACTGCCCAC 719

Qy        449 CCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCT 508
              | | | | | ||  |    |  | | | | | |  |  | ||||| ||  || ||
Db        720 CGACCGGCCGCGCCCCGCCGTCGCGCGACGGCGCCGGTTCCGCCGTCGCCTTCGACGTGCC 779

Qy        509 CTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCT 568
              | ||||   |  |||  |  | |  | |  |||  |  | |    || | |
Db        780 CGCCCCGCTGGGCACCCGGCTGCGCGAGCTGGCCGCCGCGCACGACAGCACCCCGTTCAT 839

Qy        569 CCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGT 628
                ||  ||  |  ||       || ||   ||  | || | | ||||  | ||  | |||
```

```
Db        840 GCTGTTCCTCGCCGCCTACCAGAGCCTGCTGGCCCGCCACACCGGCCGCGCGGACATCGC 899

Qy        629 CGGCCACCTCCAG---------CGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCG 679

Db        900 CGTCGGCACCGTGGTCTCCGGCCGCGGCCGCCCCGAGCTGCAACGGCTCATCGGCTACGG 959

Qy        680 CCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCT 739

Db        960 CATCAACACCCTGGTGATGCGCGGCGACCTCTCCGGCGACCCCGCCTTCACCGACCTGCT 1019

Qy        740 CGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCT 799

Db       1020 GGCCCGCACCAGGGCCACGGTCCTGGACGCCTACGACCACCAGGACGTCGCCTTCGCCCA 1079

Qy        800 GCACTGGGGCGC 811

Db       1080 GGTCGTCGACGC 1091


RESULT 8
US-10-425-115-156650
; Sequence 156650, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 156650
;   LENGTH: 1583
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_74443C.1
US-10-425-115-156650

  Query Match             9.8%;  Score 88.6;  DB 8;  Length 1583;
  Best Local Similarity   46.7%;  Pred. No. 1.1e-11;
  Matches  406;  Conservative    0;  Mismatches  439;  Indels   24;  Gaps    3;

Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65

Db        326 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 385

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125

Db        386 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 445

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185

Db        446 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 502

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245

Db        503 CTGCTGAGCCAGCTGGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 562

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305

Db        563 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 622

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365

Db        623 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC------------C 670

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425

Db        671 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 730

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485

Db        731 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 781

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545

Db        782 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 841

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605

Db        842 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 901

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665

Db        902 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCGGCTTCGACGTGCTC 961

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725

Db        962 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCCCGCAAGCCCCTC 1021

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785

Db       1022 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCGCACGCGCACGGCGCCGTG 1081

Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845

Db       1082 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1141

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874

Db       1142 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1170


RESULT 9
US-09-940-316B-1/c
; Sequence 1, Application US/09940316B
; Publication No. US20030175901A1
; GENERAL INFORMATION:
;  APPLICANT: KOSAN BIOSCIENCES, Inc.
;  APPLICANT:  REEVES, CHRISTOPHER
;  APPLICANT:  CHU, DANIEL
;  APPLICANT:  KHOSLA, CHAITAN
;  APPLICANT:  SANTI, DANIEL
;  APPLICANT:  WU, KAI
;  TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTHA
;  TITLE OF INVENTION:  GENE CLUSTER
;  FILE REFERENCE: 30062-20026.11
;  CURRENT APPLICATION NUMBER: US/09/940,316B
;  CURRENT FILING DATE:  2001-08-27
;  PRIOR APPLICATION NUMBER: 09/410,551
;  PRIOR FILING DATE: 1999-10-01
;  PRIOR APPLICATION NUMBER: US 60/139,650
;  PRIOR FILING DATE: 1999-06-17
;  PRIOR APPLICATION NUMBER: US 60/123,810
;  PRIOR FILING DATE: 1999-03-11
;  PRIOR APPLICATION NUMBER: US 60/102,748
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 77536
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (52275)...(71465)
US-09-940-316B-1

  Query Match             9.8%;  Score 88.4;  DB 3;  Length 77536;
  Best Local Similarity   45.5%;  Pred. No. 6.6e-12;
  Matches  314;  Conservative    0;  Mismatches  376;  Indels    0;  Gaps    0;
```

```
Qy        124 ACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGT 183
Db      44340 ACGGCCTCGTCGACGCGGGCGAGCTGCGACCCGGCCAGAAGGTCCTGGTCCACGCCGCGA 44281

Qy        184 CCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACG 243
Db      44280 CGGGCGGTGTCGGCATGGCGGCGACCCGGATCGCCCGCCACCTCGGCGCCGAGGTGTACG 44221

Qy        244 TCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGC 303
Db      44220 CCACCGCCAGTCCCGCCAAACAACACCTGCTGCACGCCGACGGCTTCGACACCGACCATG 44161

Qy        304 GCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCC 363
Db      44160 TGGCGAACTCCCGCAGCGCCGCGTTCGCCGACACCTTCCCGCCGGTCGACGTCGTGCTCA 44101

Qy        364 TCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGA 423
Db      44100 ACTCGCTCACCGGTGAACTCCTCGACGCGTCCATCGGCCTGCTCGCACCGGGCGGCCGGT 44041

Qy        424 TCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACG 483
Db      44040 TCGTCGAGATGGGCAAGACCGACATCCGCCACGCCGCCCAGCAGCCGTTCGACCTGGCCG 43981

Qy        484 ATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGA 543
Db      43980 ACGTGGATCCCGCGCGCCTGCGGGAGATCCTCGAACTGCTCCTCGACCTGTTCGACCGCG 43921

Qy        544 CCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCA 603
Db      43920 GTGAGCTGTGCCCGCTGCCGGTGCAGCCGTGGGACATCCGCCGTGCGCGGGACGCGTTCG 43861

Qy        604 GCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCG 663
Db      43860 CGTGGATGAGCCACGCCCGGCACACCGGCAAGATGGTCCTCACCGTGCCGCCTCCGATCG 43801

Qy        664 GGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCA 723
Db      43800 GCCCGGACGCGCCGGTCCTGGTCACCGGGGAGGACGCCGACACCGTCGCCGGCCTGCTCC 43741

Qy        724 CCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCG 783
Db      43740 GCGACGAGACCGGCCTCACCCACGTGACCACCGCAGCCGAGGAAGCCGCGAACCCCGGCG 43681

Qy        784 AGCTCCGCCTCGTCCTGCACTGGGGCGCGG 813
Db      43680 CGCTCGTGCACGTCCTGCTCGACGGCCCGG 43651


RESULT 10
US-10-760-493-36
; Sequence 36, Application US/10760493
; Publication No. US20050187167A1
; GENERAL INFORMATION:
;  APPLICANT: Ecopia BioSciences Inc
;  APPLICANT:  Farnet, Chris M.
;  APPLICANT:  McAlpine, James B.
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Bachmann, Brian O.
;  APPLICANT:  Piraee, Mahmood
;  TITLE OF INVENTION: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
;  TITLE OF INVENTION:  PHARMACEUTICALS
;  FILE REFERENCE: 3004-9US
;  CURRENT APPLICATION NUMBER: US/10/760,493
;  CURRENT FILING DATE:  2004-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/441,123
;  PRIOR FILING DATE: 2003-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/469,810
;  PRIOR FILING DATE: 2003-05-13
;  PRIOR APPLICATION NUMBER: USSN 60/491,516
;  PRIOR FILING DATE: 2003-08-01
;  PRIOR APPLICATION NUMBER: USSN 60/494,568
;  PRIOR FILING DATE: 2003-08-13
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
;   LENGTH: 22533
;   TYPE: DNA
;   ORGANISM: Streptomyces aizunensis
US-10-760-493-36

  Query Match             9.7%;  Score 87.8;  DB 9;  Length 22533;
  Best Local Similarity   44.4%;  Pred. No. 1.1e-11;
  Matches  399;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
Db      10554 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 10613

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
Db      10614 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 10673

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
Db      10674 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 10733

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
Db      10734 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 10793

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
Db      10794 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 10853

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
Db      10854 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 10913

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
Db      10914 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 10973

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
Db      10974 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 11033

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
Db      11034 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 11093

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
Db      11094 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 11153

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
Db      11154 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 11213

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
Db      11214 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 11273

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
Db      11274 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 11333

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
Db      11334 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 11393

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
Db      11394 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 11452
```

```
RESULT 11
US-10-760-493-18
; Sequence 18, Application US/10760493
; Publication No. US20050187167A1
; GENERAL INFORMATION:
;  APPLICANT: Ecopia BioSciences Inc
;  APPLICANT:  Farnet, Chris M.
;  APPLICANT:  McAlpine, James B.
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Bachmann, Brian O.
;  APPLICANT:  Piraee, Mahmood
;  TITLE OF INVENTION: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
;  TITLE OF INVENTION:  PHARMACEUTICALS
;  FILE REFERENCE: 3004-9US
;  CURRENT APPLICATION NUMBER: US/10/760,493
;  CURRENT FILING DATE:  2004-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/441,123
;  PRIOR FILING DATE: 2003-01-21
;  PRIOR APPLICATION NUMBER: USSN 60/469,810
;  PRIOR FILING DATE: 2003-05-13
;  PRIOR APPLICATION NUMBER: USSN 60/491,516
;  PRIOR FILING DATE: 2003-08-01
;  PRIOR APPLICATION NUMBER: USSN 60/494,568
;  PRIOR FILING DATE: 2003-08-13
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 164051
;   TYPE: DNA
;   ORGANISM: Streptomyces aizunensis
US-10-760-493-18

  Query Match             9.7%;  Score 87.8;  DB 9;  Length 164051;
  Best Local Similarity   44.4%;  Pred. No. 8.1e-12;
  Matches  399;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||   |||||  |    | | |    |  |||||||| | |||  | ||    |||   |
Db     114342 CGCATGGGACCTGCGGCAGGCCCCCGAGGCGTTCCGCTACCTCAGCCAGGCACGCCACGT 114401

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              |||   |||||    |   | | ||| |||| ||||   |||| ||  |  ||
Db     114402 CGGCAAGATCGTGCTCACCCTGCCGGGCGAGTGGAACTCGCAGGGCACCGTCCTCATCAC 114461

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              || ||    || |  |    ||   | | |  | ||     ||      | || |  | |
Db     114462 CGGCGGCACCGGCACCCTCGGCGCGGTGGTCGCCCGGCACGCCGTCACCACCCGCGGCGC 114521

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
                || || |  |||||  |   | || |  ||  | || |     |  |||    | |
Db     114522 CCGCCGCCTGCTGCTCACCAGTCGGCGCGGCGAGGCCGCCGCCGGCGCCGCCGAACTCGC 114581

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
              || ||   |  |  |   ||   ||  |          |  ||    || || || || ||
Db     114582 CGCCGAACTGCGGGAACTGGGCGCCGAGGTCACGATCGCGGCCTGCGACGCCGCCGACCG 114641

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
                || ||||  |  || |   ||  ||      |         |  |  |  |  |  | |
Db     114642 CGACGCGCTCGCCGCGCTCATCGAATCCATACCGTCAGAGCACCCGCTGACGGCCGTCAT 114701

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
              || |  |  |   |    ||||| || || || |||      |  | ||| ||| | | |
Db     114702 CCACACCGCCGGAGTCCTCGACGACGGCGTCGTCGACTCGCTGACCCCCGAGCGCCTGTC 114761

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGA 481
              | | | |   || |   | | |  |  |   ||  | ||| | || |||
Db     114762 CACGGTCCTGCGCCCGAAGGTGGACGCCGCCTGGAACCTGCACGAGCTGACCCGTCACCT 114821

Qy        482 CGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCC 541
              ||| ||   |  |  | ||||     || | | ||||        |  |  || |  |
Db     114822 CGACCTGGCCGACTTCGTCCTGTTCTCCTCCGCCGCCGGCACCTTCGGCGGCGCCGGACA 114881

Qy        542 GACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGAT 601
              | || |   |  |  | |  |   |  |||| || | |||  ||       |||
Db     114882 GGCCAACTACGCGGCCGCGAACGTCTTCCTGGACGCCCTCGCCCGCCACCGGCACGCCCA 114941

Qy        602 CAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCT 661
              | ||    || |  | | |    |  ||||          |  | || || ||  |
Db     114942 CGGCCTCGCCGCCACCTCCCTGGCCTGGGGCCTGTGGGCCGAGGCCAGCGGCATGACCGG 115001

Qy        662 CGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCC 721
              ||        | | ||  |         |  ||   ||  |||   |||||   |||   |
Db     115002 CGAACTCGACACCGCCGACAAGGACCGGATGACGCGCTCCGGCGTCCTCGGCCTCTCCTC 115061

Qy        722 CACCATGACGCGCGC---GCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
              |     |      ||   ||||||| | || ||  |   | |||||||  ||  || |||
Db     115062 CGAAGAGGGCGTGGCGCTGCTCGACACCGCACGGCTCACCGGCGACGCCCTCCTCGTCCC 115121

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGAT 838
              |    | |||  ||| | | ||   |  |  | |     |  |  |  | || | || |
Db     115122 CATGCACCTCGACCTGGCGCCGCTGCGCCGGACCGACGCCAGCATGGTCCCCGCCCTGCT 115181

Qy        839 CGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCGCGCCG 897
              || |   ||||| |   | | ||| |  |  |  |   ||   |  |   ||||||
Db     115182 GCGCGGCCTGGTCCGCGCCCCCGCCCGCAGGGCCGTCGGAGCCACCGCCGCCGGCGCCG 115240


RESULT 12
US-10-282-122A-25736
; Sequence 25736, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25736
;   LENGTH: 3948
;   TYPE: DNA
;   ORGANISM: Mycobacterium avium
US-10-282-122A-25736
```

```
  Query Match             9.7%;  Score 87.4;  DB 7;  Length 3948;
  Best Local Similarity   45.9%;  Pred. No. 1.8e-11;
  Matches  390;  Conservative    0;  Mismatches  441;  Indels   18;  Gaps    2;

Qy         43 TCTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGC 102
Db        221 TCTGTGAGCGCTGCGGCGTCGAGGTGACCCGCGCCAAGGTGCGCCGCGAGCGCATGGGCC 280

Qy        103 AGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGC 162
Db        281 ACATCGAGTTGGCCGCGCCCGTCACGCACATCTGGTACTTCAAGGGCGTGCCGTCGCGAT 340

Qy        163 CCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCC 222
Db        341 TGGGCTACCTGCTCGACCTGGCGCCCAAGGACCTCGAGAAGATCATCTACTTCGCCGCGT 400

Qy        223 GGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGG 282
Db        401 ACGTGATCACGTCCGTCGACGAGGAGATGCGGCACAACGAGCTCTCCACCCTCGAGGCCG 460

Qy        283 ATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCG 342
Db        461 AGATGATGGTGGAGCGCAAGGCCGTCGAGGACCAGCGCGACGCCGACCTGGAGGCGCGCG 520

Qy        343 GCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCG 402
Db        521 CCCAGAAGCTGGAGGCCGACCTGGCCGAGCTGGAGGCCGAGGGCGCCAAGGCCGACGCGC 580

Qy        403 AGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACC 462
Db        581 GGCGCAAGGTCCGCGACAGCGGCGAGCGGGAGATGCGCCAGATCCGCGACCGCGCCCAGC 640

Qy        463 TCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGC 522
Db        641 GCGAGCTGGACCGGCTGGAGGACATCTGGAACACCTTCACCAAGCTGGCCCCCAAGCAGC 700

Qy        523 GGCGGGAGGAGGCG------------GAGCCGACCGATCGGGTCGAGGTGGCGATGCTCC 570
Db        701 TGATCGTCGACGAGAACCTCTACCGCGAGCTGGTCGACCGTTACGGCGAGTACTTCACCG 760

Qy        571 TCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCG 630
Db        761 GCGCCATGGGCGCGGAGTCGATCCAGAAGCTCATCGAGAACTTCGACATCGACGCCGAGG 820

Qy        631 GCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGG 690
Db        821 CCGAGCAGCTGCGTGACGTCATCCGAAACGGCAAGGGGCAGAAGAAGCTTCGCGCCCTCA 880

Qy        691 GGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCC 750
Db        881 AGCGGCTCAAGGTGGTCGCCGCGTTCCAACAGAGCGGCAACTCGCCGAT------GGGCA 934

Qy        751 GCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCG 810
Db        935 TGGTGCTCGACGCCGTCCCGGTGATCCCGCCGGAGCTGCGCCCGATGGTGCAGCTCGACG 994

Qy        811 CGGTGCGGATGAGCGATCACGACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCG 870
Db        995 GTGGCCGGTTCGCCACCAGCGACCTCAACGACCTGTACCGCCGGGTGATCAACCGCAACA 1054

Qy        871 AGGCGGTCA 879
Db       1055 ACCGGCTCA 1063


RESULT 13
US-10-425-114-15723
; Sequence 15723, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15723
;   LENGTH: 1254
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3060-061-C11_FLI
US-10-425-114-15723

  Query Match             9.7%;  Score 87;  DB 7;  Length 1254;
  Best Local Similarity   46.6%;  Pred. No. 2.6e-11;
  Matches  405;  Conservative    0;  Mismatches  440;  Indels   24;  Gaps    3;

Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
Db        279 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 338

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
Db        339 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 398

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
Db        399 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 455

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
Db        456 CTGCTGAGCCAGCTGGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 515

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
Db        516 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 575

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
Db        576 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC------------C 623

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
Db        624 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCATGGCGTTC 683

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
Db        684 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 734

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
Db        735 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 794

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
Db        795 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 854

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
Db        855 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 914

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
Db        915 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCGCGCGAAGCCACTC 974

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
Db        975 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCACACGCGCACGGCGCCGTG 1034
```

```
Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845
              ||| |||     |   | |||  ||| | ||  |     | | |||| |   ||| |    |
Db       1035 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1094

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874
              | || |     ||   | | |  | ||||
Db       1095 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1123


RESULT 14
US-10-425-114-15784
; Sequence 15784, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15784
;   LENGTH: 1434
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3060-077-D5_FLI
US-10-425-114-15784

  Query Match             9.7%;  Score 87;  DB 7;  Length 1434;
  Best Local Similarity   46.6%;  Pred. No. 2.5e-11;
  Matches  405;  Conservative    0;  Mismatches  440;  Indels   24;  Gaps    3;

Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
              ||   | | |  |||||  | | || |  ||  |||||          ||  |  |
Db        269 CGCAGCACCGTGGACGTGGAGCGCCTGGGCCCGGAGCTGCAGCGCATGCGCGCGGGCCTC 328

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
               |   | | |||      || |  | || |  || |  || || |  | | | | ||  | |
Db        329 ATCCGCCTCTGCGCCGACGCCGAGGCGCTGCTGGAGGCGCACTACTCCGACCTGCTGGCC 388

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
              ||  ||| |  | |||   ||   ||||||| ||  |||| | |  |  | | |||    |
Db        389 GCCTTCGACAACCCGCTGGACCACCTGCCGCTCTTCCCCTACTTCACCAACTACCT---C 445

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              || | |   |   | ||||   ||   |   |||  |||  |    || || |  |
Db        446 CTGCTGAGCCAGCTGGAGCACGGCCTGCTGGCGCGCCACGTCCCGGGCCCGCCCCCCTCC 505

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
              | ||| |  ||  |   || | |||| ||          | ||||  || |  ||||
Db        506 CGCGTCGCCTTCGTCGGCTCCGGCCCGCTCCCGCTCAGCTCCCTCGTGCTCGCCTCGCGC 565

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
              | |||  |      | |   || | || | || ||| ||| ||  |             |
Db        566 CACCTCCCCGCCGCCGCCTTCGACAACTACGACATCTGCGGGGACGC-----------C 613

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              || ||| ||||  |  | |||    || ||| | || |  |||||||||   |||| ||
Db        614 AACGACCGCGCGCGCCGCCTCGTGCGCGCCGACGCCGCGCTCGCCGCGCGCGCATGGCGTTC 673

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
                  | ||||  || |   |  |  || |  ||  || | |  |   ||  ||||
Db        674 ---------CGCACCTCCGACGTCGCCCACGTCACCCGGGAACTGGCCGCCTACGACGTC 724

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              ||  ||| ||    |||| |  | |  | |||| |  |  || |   |  |    |  |

Db        725 GTCTTCCTGGCGGCGCTCGTGGGCATGGCCGCCGAGGAGAAGGCGCGCGTCGTCGAGCAC 784

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
                ||    | |||||      |  |       | ||||||    | |||        ||
Db        785 CTCGGGAGGCACATGGCGCCCGGCGCCGCGCTGGTGGTGCGGAGCGCGCACGGCGCGCGC 844

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
                ||    || |      | | |  | |        || | | || || |   | |
Db        845 GGCTTCCTGTACCCCGTGGTTGACCCCGAGGAGATCCGCCGCGGCGGCTTCGACGTGCTC 904

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
              |   | ||| ||       | ||| ||||  |  ||    | |||||  | | ||   |
Db        905 GCCGTGCACCACCCGGAGGGGGAGGTGATCAACTCCGTCATCATCGCGCGCGAAGCCACTC 964

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
              |      ||      |||| | |  || | |  |  ||||   ||   ||  ||||| |
Db        965 CCGGTCGTCGACGAGCACGCCGTCGCCGGGGTGGGGCACGCACACGCGCACGGCGCCGTG 1024

Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGG 845
              ||| |||     |   | |||  ||| | ||  |     | | |||| |   ||| |    |
Db       1025 CTCAGCCGGCCCTGCCTCTGCTGCGAGATGGAGGCCCGGGCTCACCAGAAGATGGAGGAG 1084

Qy        846 GAGGTCCACCGGCTCTTCCGGATCGAGGC 874
              | || |     ||   | | |  | ||||
Db       1085 GTGGCCATGGAGCAGCTGCCGTCCTAGGC 1113


RESULT 15
US-10-203-295-2
; Sequence 2, Application US/10203295
; Publication No. US20040115762A1
; GENERAL INFORMATION:
;  APPLICANT: Zotchev, Sergey Borisovich
;  APPLICANT:  Sekurova, Olga Nikalayivna
;  APPLICANT:  Fjaervik, Epsen
;  APPLICANT:  Brautaset, Trygve
;  APPLICANT:  Strom, Arne Reidar
;  APPLICANT:  Valla, Svein
;  APPLICANT:  Ellingsen, Trond Erling
;  APPLICANT:  Sletta, Havard
;  APPLICANT:  Gulliksen, Ole-Martin
;  TITLE OF INVENTION: Novel genes encoding a nystatin polyketide synthase and their
;  TITLE OF INVENTION:  manipulation and utility
;  FILE REFERENCE: 1181-265
;  CURRENT APPLICATION NUMBER: US/10/203,295
;  CURRENT FILING DATE:  2003-05-19
;  PRIOR APPLICATION NUMBER: PCT/GB 01/00509
;  PRIOR FILING DATE: 2001-02-08
;  PRIOR APPLICATION NUMBER: GB 0002840.7
;  PRIOR FILING DATE: 2000-02-08
;  PRIOR APPLICATION NUMBER: GB 0008786.6
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: GB 0009387.2
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 27541
;   TYPE: DNA
;   ORGANISM: Streptomyces noursei ATCC 11455
US-10-203-295-2

  Query Match             9.6%;  Score 86.8;  DB 7;  Length 27541;
  Best Local Similarity   45.2%;  Pred. No. 1.8e-11;
  Matches  362;  Conservative    0;  Mismatches  432;  Indels    6;  Gaps    1;

Qy         52 GCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCC 111
              ||| | | |  |   || ||  ||||     | ||| ||  | |   | |||  |
Db       7767 GCGTCGTACCGGTCCTGCTCTCCGGCCGCACCCCGGACGCCCTGCGCGGCCAGGCCGCGG 7826

Qy        112 GCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACC 171
              || ||    || |   | |  |||| ||| | ||  ||||     |    | | | ||
Db       7827 CCCTGCGCGCCGCCCTCGACACCGGCCGGCGGCCCGACCTGCTCGACCTCGCACACTCCC 7886
```

```
Qy        172 TCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCG 231
              ||  |  | |   |  | ||        |    |  ||| |  || |  |     || ||
Db       7887 TCGCCACCACCCGCGCCGGGTTCGAGCACCGCGCCGTCCTCCTCGCCACCGACCACCCCG 7946

Qy        232 CGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCG 291
              |  ||  | ||| | ||| || || ||  ||  ||      |   |   | | || |
Db       7947 CCCTGACCGACGGCCTCACCGCCCTCGCCGACGCCGACGACCCGGCCGCCGCCCCCGCCT 8006

Qy        292 TCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGG 351
              |  |     |   |    | ||  || |  ||||| |  |   | |||| |||  |
Db       8007 GGATCACCGGCACCACCCGGGCCGAGACCCGGCTCGCCGTCCTGTTCACCGGCCAGGGCG 8066

Qy        352 TGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCG 411
              ||||  ||     |  |    || || |     |   | |  ||    ||||
Db       8067 CCCAACGCCTCGGCGCGGGACGGGAACTCGCCGCCCGTTTCCCGGCGTTCGCCACCGCCC 8126

Qy        412 CGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGA 471
              |||| ||| ||| |   || | | |  |  || |  |   | |   |  | |
Db       8127 TCGACGCGGCGCTCGACGCCTTCACCCCGCACCTCGACCGCCCCCTGCGCGAGGTCCTGT 8186

Qy        472 TCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGG 531
              ||   ||||    ||  |||| || |  ||     |  || | | || |     |  |
Db       8187 GGGGCACCGACGCCGCCCTGCTCGACCGCACCGCATACGCCCAGCCGGCCCTCTTCGCCG 8246

Qy        532 AGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGA 591
              |  | | ||  | | ||| |    |   || |   | ||  |  |  |   | |
Db       8247 TCGAAGTGGCGCTCTACCGGCTGATCGAATCGTTCGGCGTCCGCCCCGACCACCTCGCCG 8306

Qy        592 CCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCC 651
              ||||    || || |||||| || | ||| | ||| |||   ||||   | |  |  |
Db       8307 GCCACTCCGTCGGCGAGATCGTCGCCGCGCACCTCGCCGGGGTCCTCTCCCTGGCCGACG 8366

Qy        652 GCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCG 711
              ||    |||||   |  |   |  |||||  || ||| | |   ||||||      ||
Db       8367 CCGCCACCCTCGTCGCCGCCCGCGGTCGCCTGATGCAGGCGCTGCCCGACGG------CG 8420

Qy        712 CGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCG 771
              ||        | || |   |    ||   | ||  | |  |  ||| | || || |  ||
Db       8421 GGGCGATGATCGCCGTCCAGGCGTCGGAAGCCGACGTCGCCCCGCTGCTCGCCGGGCACG 8480

Qy        772 ACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACG 831
              | ||||  | || | ||  |  ||||     |     ||| ||     ||  ||    ||
Db       8481 AGGACCAGGTCGCGATCGCCGCCGTCAACGGCCCCTCCGCCGTCGTCCTGTCCGGCGCCG 8540

Qy        832 ACGTGATCGGCAGGGAGGTC 851
              | |  | || ||  | | ||
Db       8541 AAGCCACCGTCACCGCGCTC 8560
```

Search completed: April 27, 2006, 14:39:58
Job time : 911.317 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 565.78 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_75000_75900
Perfect score: 901
Sequence: 1 acggccgggacgcggacgac.....gaggaggatcgcgccgcgga 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	87	9.7	9222	7	US-10-496-351-2	Sequence 2, Appli
c	2	87	9.7	82746	7	US-10-496-351-56	Sequence 56, Appl
	3	83.6	9.3	1250	11	US-11-096-568A-23615	Sequence 23615, A
	4	82.2	9.1	113193	7	US-10-478-943D-1	Sequence 1, Appli
	5	82	9.1	4770	7	US-10-496-351-3	Sequence 3, Appli
	6	81	9.0	14055	7	US-10-496-351-1	Sequence 1, Appli
	7	79.6	8.8	9975	7	US-10-496-351-4	Sequence 4, Appli
	8	78.6	8.7	1026	7	US-10-496-351-17	Sequence 17, Appl
	9	77.8	8.6	2247	7	US-10-496-351-8	Sequence 8, Appli
	10	77.4	8.6	88421	14	US-11-205-109-1	Sequence 1, Appli
	11	77.2	8.6	3354	7	US-10-496-351-12	Sequence 12, Appl
	12	76.2	8.5	2736	8	US-10-858-730-38	Sequence 38, Appl
c	13	75.8	8.4	88421	14	US-11-205-109-1	Sequence 1, Appli
	14	74.8	8.3	11070	14	US-11-075-185-34	Sequence 34, Appl
	15	74.8	8.3	78869	14	US-11-075-185-1	Sequence 1, Appli
c	16	74.2	8.2	9222	7	US-10-496-351-2	Sequence 2, Appli
	17	74.2	8.2	82746	7	US-10-496-351-56	Sequence 56, Appl
	18	73.2	8.1	2742	7	US-10-496-351-22	Sequence 22, Appl
	19	73.2	8.1	10968	14	US-11-075-185-35	Sequence 35, Appl
c	20	73.2	8.1	37507	8	US-10-522-037-2	Sequence 2, Appli
	21	72.4	8.0	1695	11	US-11-096-568A-12871	Sequence 12871, A
c	22	72	8.0	14055	7	US-10-496-351-1	Sequence 1, Appli
	23	71.8	8.0	1546	11	US-11-096-568A-23918	Sequence 23918, A
	24	71.4	7.9	1583	11	US-11-096-568A-12501	Sequence 12501, A
	25	71	7.9	1029	7	US-10-496-351-51	Sequence 51, Appl
	26	71	7.9	11905	7	US-10-496-351-57	Sequence 57, Appl
c	27	70.6	7.8	1042	11	US-11-096-568A-19268	Sequence 19268, A
	28	70.2	7.8	2489	11	US-11-096-568A-14645	Sequence 14645, A
c	29	70	7.8	6264	14	US-11-075-185-58	Sequence 58, Appl
c	30	70	7.8	78869	14	US-11-075-185-1	Sequence 1, Appli
	31	69.6	7.7	3305	14	US-11-017-550-68	Sequence 68, Appl
	32	69.6	7.7	12391	14	US-11-017-550-67	Sequence 67, Appl
	33	69.2	7.7	1173	11	US-11-096-568A-18583	Sequence 18583, A
	34	68.8	7.6	1504	11	US-11-096-568A-26094	Sequence 26094, A
	35	68.6	7.6	1110	8	US-10-432-483-14	Sequence 14, Appl
	36	68.6	7.6	8651	8	US-10-432-483-48	Sequence 48, Appl
	37	68.4	7.6	1590	8	US-10-858-730-162	Sequence 162, App
	38	68	7.5	1368	11	US-11-228-659-72	Sequence 72, Appl
	39	68	7.5	1573	11	US-11-096-568A-24540	Sequence 24540, A
c	40	68	7.5	2742	7	US-10-496-351-22	Sequence 22, Appl
	41	68	7.5	7608	11	US-11-228-659-60	Sequence 60, Appl
	42	67.8	7.5	718	11	US-11-096-568A-9054	Sequence 9054, Ap
	43	67.8	7.5	1164	7	US-10-496-351-48	Sequence 48, Appl
	44	67.8	7.5	1536	11	US-11-096-568A-20769	Sequence 20769, A
	45	67.6	7.5	1171	11	US-11-096-568A-21593	Sequence 21593, A

ALIGNMENTS

RESULT 1
US-10-496-351-2
; Sequence 2, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 9222
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-2

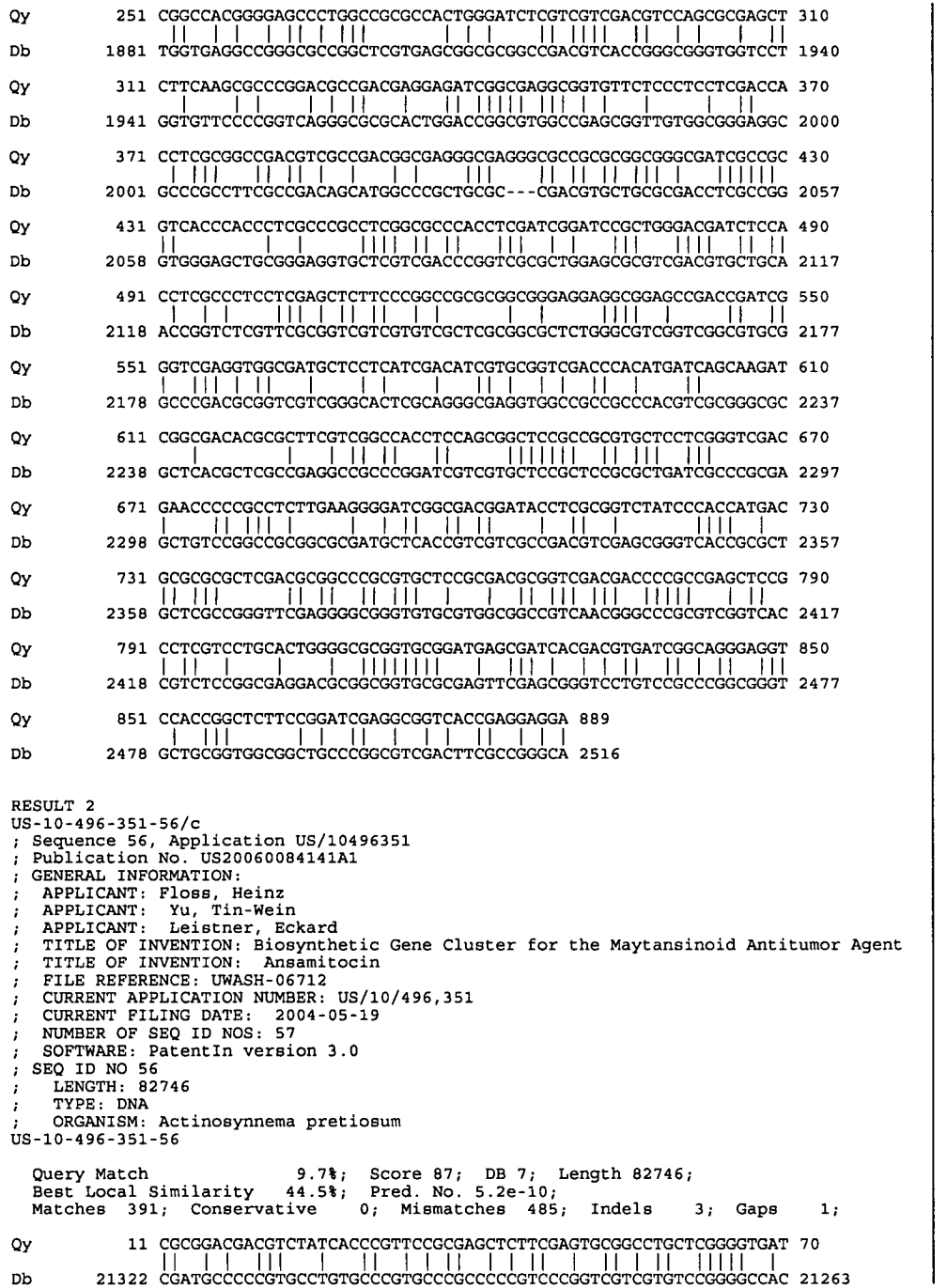
Query Match 9.7%; Score 87; DB 7; Length 9222;
Best Local Similarity 44.5%; Pred. No. 8.2e-10;
Matches 391; Conservative 0; Mismatches 485; Indels 3; Gaps 1;

```
Qy          11 CGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGAT 70
               ||  | |  |||    |   ||  | | ||  | | ||  |  || | ||  |||||  |
Db        1641 CGATGCCCCCGTGCCTGTGCCCGTGCCCGCCCCCGTCCCGGTCGTCGTGTCCGGGGCCAC 1700

Qy          71 CGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGT 130
               | || || || |   ||| ||  |  |   | |  |   | || || | |      | |
Db        1701 CACCGCCGGTCTGCGCGCCCAGGCGGCGCGCCTGGCCGGGCACCTGCGGGAGCGGCCCGC 1760

Qy         131 CGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCG 190
                |||| |     |||     | ||    ||   | || |  ||| |    |  ||
Db        1761 GCTCGGCCCCGAGGACGTCGCCCGGCCGCTGCTGCTCTCCCGCGCCCAGCGCGAGCGGCG 1820

Qy         191 GGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCAT 250
               ||  || | |   | | |  | | |  |  | | || ||| | |   |  | |||
Db        1821 CGCCGTCGTGGTCGCCCGCGACCGGGAGGCGCTGCTCGGCGGGCTCGACGCCCTCGCGGG 1880
```



```
Qy        251 CGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCT 310
Db       1881 TGGTGAGGCCGGGCGCCGGCTCGTGAGCGGCGCGGCCGACGTCACCGGGCGGGTGGTCCT 1940

Qy        311 CTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCA 370
Db       1941 GGTGTTCCCCGGTCAGGGCGCGCACTGGACCGGCGTGGCCGAGCGGTTGTGGCGGGAGGC 2000

Qy        371 CCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGC 430
Db       2001 GCCCGCCTTCGCCGACAGCATGGCCCGCTGCGC---CGACGTGCTGCGCGACCTCGCCGG 2057

Qy        431 GTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCA 490
Db       2058 GTGGGAGCTGCGGGAGGTGCTCGTCGACCCGGTCGCGCTGGAGCGCGTCGACGTGCTGCA 2117

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550
Db       2118 ACCGGTCTCGTTCGCGGTCGTCGTGTCGCTCGCGGCGCTCTGGGCGTCGGTCGGCGTGCG 2177

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610
Db       2178 GCCCGACGCGGTCGTCGGGCACTCGCAGGGCGAGGTGGCCGCCGCCCACGTCGCGGGCGC 2237

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGAC 670
Db       2238 GCTCACGCTCGCCGAGGCCGCCCGGATCGTCGTGCTCCGCTCCGCGCTGATCGCCCGCGA 2297

Qy        671 GAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGAC 730
Db       2298 GCTGTCCGGCCGCGGCGCGATGCTCACCGTCGTCGCCGACGTCGAGCGGGTCACCGCGCT 2357

Qy        731 GCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCG 790
Db       2358 GCTCGCCGGGTTCGAGGGGCGGGTGTGCGTGGCGGCCGTCAACGGGCCCGCGTCGGTCAC 2417

Qy        791 CCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGGGAGGT 850
Db       2418 CGTCTCCGGCGAGGACGCGGCGGTGCGCGAGTTCGAGCGGGTCCTGTCCGCCCGGCGGGT 2477

Qy        851 CCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGA 889
Db       2478 GCTGCGGTGGCGGCTGCCCGGCGTCGACTTCGCCGGGCA 2516
```

RESULT 2
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 82746
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

Query Match 9.7%; Score 87; DB 7; Length 82746;
Best Local Similarity 44.5%; Pred. No. 5.2e-10;
Matches 391; Conservative 0; Mismatches 485; Indels 3; Gaps 1;

```
Qy         11 CGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGAT 70
Db      21322 CGATGCCCCCGTGCCTGTGCCCGTGCCCGCCCCCGTCCCGGTCGTCGTGTCCGGGGCCAC 21263

Qy         71 CGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGT 130
Db      21262 CACCGCCGGTCTGCGCGCCCAGGCGGCGCGCCTGGCCGGGCACCTGCGGGAGCGGCCCGC 21203

Qy        131 CGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCG 190
Db      21202 GCTCGGCCCCGAGGACGTCGCCCGGCCGCTGCTGCTCTCCCGCGCCCAGCGCGAGCGGCG 21143

Qy        191 GGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCAT 250
Db      21142 CGCCGTCGTGGTCGCCCGCGACCGGGAGGCGCTGCTCGGCGGGCTCGACGCCCTCGCGGG 21083

Qy        251 CGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCT 310
Db      21082 TGGTGAGGCCGGGCGCCGGCTCGTGAGCGGCGCGGCCGACGTCACCGGGCGGGTGGTCCT 21023

Qy        311 CTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCA 370
Db      21022 GGTGTTCCCCGGTCAGGGCGCGCACTGGACCGGCGTGGCCGAGCGGTTGTGGCGGGAGGC 20963

Qy        371 CCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGC 430
Db      20962 GCCCGCCTTCGCCGACAGCATGGCCCGCTGCGC---CGACGTGCTGCGCGACCTCGCCGG 20906

Qy        431 GTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCA 490
Db      20905 GTGGGAGCTGCGGGAGGTGCTCGTCGACCCGGTCGCGCTGGAGCGCGTCGACGTGCTGCA 20846

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550
Db      20845 ACCGGTCTCGTTCGCGGTCGTCGTGTCGCTCGCGGCGCTCTGGGCGTCGGTCGGCGTGCG 20786

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610
Db      20785 GCCCGACGCGGTCGTCGGGCACTCGCAGGGCGAGGTGGCCGCCGCCCACGTCGCGGGCGC 20726

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGAC 670
Db      20725 GCTCACGCTCGCCGAGGCCGCCCGGATCGTCGTGCTCCGCTCCGCGCTGATCGCCCGCGA 20666

Qy        671 GAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGAC 730
Db      20665 GCTGTCCGGCCGCGGCGCGATGCTCACCGTCGTCGCCGACGTCGAGCGGGTCACCGCGCT 20606

Qy        731 GCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCG 790
Db      20605 GCTCGCCGGGTTCGAGGGGCGGGTGTGCGTGGCGGCCGTCAACGGGCCCGCGTCGGTCAC 20546

Qy        791 CCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGCAGGGAGGT 850
Db      20545 CGTCTCCGGCGAGGACGCGGCGGTGCGCGAGTTCGAGCGGGTCCTGTCCGCCCGGCGGGT 20486

Qy        851 CCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGA 889
Db      20485 GCTGCGGTGGCGGCTGCCCGGCGTCGACTTCGCCGGGCA 20447
```

RESULT 3
US-11-096-568A-23615
; Sequence 23615, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides F
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 23615
; LENGTH: 1250
; TYPE: DNA

; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1250)
; OTHER INFORMATION: Ceres Seq. ID no. 12414165
US-11-096-568A-23615

Query Match 9.3%; Score 83.6; DB 11; Length 1250;
Best Local Similarity 46.2%; Pred. No. 6.5e-09;
Matches 396; Conservative 0; Mismatches 449; Indels 13; Gaps 3;

```
Qy         14 GGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGG 73
Db         63 GGTCGTCGTGGACGAGTGCCGGGGCGTGCTGTCCGTGTACAGCGACGGCACGGTGGTGCG 122

Qy         74 CGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGT 133
Db        123 GCGCGCGCAGCCGGGGTTCGCGACGCCGGTGCGGGACGACGGCACGGTCGACTGGAAGGA 182

Qy        134 CGGCTCGCGCCACGCGCTGCCGCG----CTCGCCCTATTACCTCCTCCACCCGTCCCTCC 189
Db        183 CGTGACGTTCGACGAGGCCCCGCGGGCTGGCGCTCCGCCTCTACCTCCCCCGGGACCGCG 242

Qy        190 GGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCA 249
Db        243 GGGCCGCGGCCGGGCGGCGCCTGCCGGTCTTCTTCTACTACCACGGCGGCGGCTTCTGCA 302

Qy        250 TCG------GCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGC 303
Db        303 TCGGGTCCCGCGCCTGGCCCAACTGCCAGAACTACTGCCTGCGCCTGGCCTCCGACCTGG 362

Qy        304 GCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCC 363
Db        363 GCGCGCTCGTGGTGGCGCCGGACTACCGCCTCGCGCCCGAGCACCGCCTCCCCGCCGCCA 422

Qy        364 TCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGA 423
Db        423 TCGACGACGGCGCGGCGGCGGTCCTGTGGCTGGCGAGGCAGGGCGGCGGGGACCCCTGGG 482

Qy        424 TCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTG---GG 480
Db        483 TCGCCGAGGCCGCCGACCTCGGCCGCGTCTTCGTGTCGGGCGACTCGGCCGGCGGCACCA 542

Qy        481 ACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGC 540
Db        543 TCGCGCACCACCTCGCCGTGCGCTTCGGTGGCTCGCCCGCGGACCTGGCCCCCGTCGCCG 602

Qy        541 CGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGA 600
Db        603 TGCGGGGATACGTCCAGCTCATGCCATTCTTCGGCGGCGTGGCGCGCACGCGGTCGGAGG 662

Qy        601 TCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCC 660
Db        663 CCGAGTGCCCCGCCGACGCTTTCCTCGACCGGCCCCTCAACGACCGCTACTGGCGCCTGT 722

Qy        661 TCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATC 720
Db        723 CGCTGCCGGAGGGCGCCACGCCCGACCACCCCGTGGCCAACCCGTTCGGCCCCGGCGCGC 782

Qy        721 CCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCG 780
Db        783 CGCCGCTGGACGCCGTCGAGTTCGCGCCCACCCTGGTCGTCGTCGGCGGCCGCGACCTCC 842

Qy        781 CCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCG 840
Db        843 TCCACGACCGCGCCGTCGACTACGCCGCCAGGCTCAGGGCCGCCGGGAAGCCCGTCGTGG 902

Qy        841 GCAGGGAGGTCCACCGGC 858
Db        903 TCAGGGACTTCCACGGCC 920
```

RESULT 4

US-10-478-943D-1
; Sequence 1, Application US/10478943D
; Publication No. US20060073574A1
; GENERAL INFORMATION:
; APPLICANT: University College Dublin, National University of Ireland
; TITLE OF INVENTION: Engineered biosynthesis of novel polyenes
; FILE REFERENCE: PC-1636US
; CURRENT APPLICATION NUMBER: US/10/478,943D
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: S2001/0527
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 113193
; TYPE: DNA
; ORGANISM: Streptomyces nodosus
US-10-478-943D-1

Query Match 9.1%; Score 82.2; DB 7; Length 113193;
Best Local Similarity 45.8%; Pred. No. 5e-09;
Matches 396; Conservative 0; Mismatches 463; Indels 6; Gaps 3;

```
Qy         34 TCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGC 93
Db       8043 TGCCCGAGCAGGGCGGCGTCCAGGTGCAGCTGTGGATCGGCAACCCGGACGCGTCCGGCC 8102

Qy         94 GGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGC 153
Db       8103 GCCGCAGCGTCACCGTCTACGGCCGGCCCGACGCCGACGAGGACGC-CCCCTGGACGTCC 8161

Qy        154 CGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCG 213
Db       8162 CACGCCACCGGTGTGCTCTCCGCCTCCCGCACCACCAGCGACTTCGACGCCACCGTCTGG 8221

Qy        214 GCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGC 273
Db       8222 CCCCCGGCCG--ACGCCGAGACCCTCCCCGTGGACGGCCTCTACGAGCGGCTCGCCGAGG 8279

Qy        274 GCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACG 333
Db       8280 GCGGCTTCGGCTACGGCCCGCTCTTCCAGGGCCTGCGGGCCGCCTGGCGCCGCGGCGACG 8339

Qy        334 AGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACG 393
Db       8340 AGGTGTTCGCCGAGGTCGTGCTGCCCGAGTCCGGCCACACCGACGCCGAGTCCTTCGGAC 8399

Qy        394 GCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCG 453
Db       8400 TGCACCCCGCGCTCCTCGACTCCGCCCTGCACGCGGCGTCGTTCGTCGACCTCGACGAGC 8459

Qy        454 GCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCC 513
Db       8460 GCGCCGCCGGCGGACTGCCCTTCTCCTGGGAGGGCG---TCAGCCTGCACGCGAGCGGTG 8516

Qy        514 CGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCA 573
Db       8517 CCACCACCCTGCGGGTGCGGCTGGCCCCGGCCGCCGGTGACGCCGTCGCGATCGCCGTCG 8576

Qy        574 TCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCC 633
Db       8577 CCGACGACTCCGGACAGCTCGTCCTGTCCGCCGACTCGCTGATCCTGCGCGCCGTCGCGG 8636

Qy        634 ACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGA 693
Db       8637 CCCGCGAGATCGACGCCGCCGCCGCGCTGGTCCGCGACGCGCTGTTCCGCCTCGACTGGG 8696

Qy        694 TCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCG 753
Db       8697 TGCCCGTCACGGCCGTCGCCGCCTCCGGCACCGCCGCGGCACTGGTCGGCGAGGACCCCT 8756

Qy        754 TGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGG 813
```

```
Db       8757 TCGGCCTGCGCGCCCTGCCCCAGTTCGGCGACCTCGCCGTCCACCCGGATCTCGCGGATC 8816

Qy        814 TGCGGATGAGCGATCACGACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGG 873
              ||        |||   ||  ||     | ||  | | | |  | ||||  | | | ||
Db       8817 TGGCCGCCGCCGACGGCGCTGTCCCGGACACCGTGCTGCTGCCGCTCACCGGCACCGGAC 8876

Qy        874 CGGTCACCGAGGAGGATCGCGCCGC 898
              | | | ||||   ||    ||||||
Db       8877 CCGACGCCGACCCGGTCACCGCCGC 8901


RESULT 5
US-10-496-351-3
; Sequence 3, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 4770
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-3

  Query Match             9.1%;  Score 82;  DB 7;  Length 4770;
  Best Local Similarity   45.8%;  Pred. No. 1.1e-08;
  Matches  360;  Conservative    0;  Mismatches  420;  Indels    6;  Gaps    2;

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172
              || || || || ||| |   ||| | |  || ||||    |    | | | |    ||||
Db       2961 CCTGCTGCCCGGCGCCGCGCTCGACCCCGGCGACGCCGACCGCGCCGCCCTCGCCGACCT 3020

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232
              |||| ||||  |  |   |        || |    | |    |||| |    | | | |
Db       3021 GCTCCGCCCGCACACCGCCGTCGCGGGCGTCCTTGCCACGCCGGGCGCCGACGTGCTCAC 3080

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292
              ||  ||| |   ||  || |   || |   ||  |||||     || |    |  ||
Db       3081 CCTGGTCCAGGCGCTCGGCGACGCGGGCGCCACCGCGCCGCTGTGGTGCGCCACCACCGG 3140

Qy        293 CGACGTCCAG---CGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGC 349
              | ||||  |     |  || |  |   |  |||||| ||| | |||    |  |  ||
Db       3141 GGCCGTCGCGATCTCCCCGCACGACGGCCCGCCGGACCCCGCCCAGGCCGCCACCTGGGG 3200

Qy        350 GGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGC 409
              ||  |  | || |||  |          ||| |  |   | |||   || |||   |||
Db       3201 ACTGGGCCGCGTCGTCGCGCTGGAGCACCCCGGCCGCTGGGGCGGGCTGGTCGACCTCGC 3260

Qy        410 CGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCG 469
              ||  |     |      |  |   | |   ||||||  | ||  ||||   | ||
Db       3261 CGACCCCGCCGACCCGCGCCAGCGCCGCGGGCTCGCCGCCGTCCTCGCCGGACCCGGCGT 3320

Qy        470 GATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGA 529
                |  | | |  ||   |||  ||||| ||| || ||    |  |  ||||||  ||
Db       3321 CGACGACGGAGGTGAGGACCAGGTCGCCATCCGCGCGCAGGGCGTGTTCGCGCGCAGGCT 3380

Qy        530 GGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTC 589
              | |  |       | |       | |   ||  |  ||   || ||  || | ||
Db       3381 CGTGCGCGCCGAACCGCAGGCCCCGGCCGCCCCGTGGACCCCGCGCGGCACGGTCCTGAT 3440

Qy        590 GACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCG 649
              |||  |   | | ||    || || | ||| | | | | ||| ||||      ||| |
Db       3441 CACCGGCGGAACCGGCGG---CGTCGCCGCGCACGTGGCCCGCCGCCTCGCCGCCCTCGG 3497

Qy        650 CCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCT 709
```

```
              |  ||  | |||||     || | || ||||  |     ||    |||||         ||
Db       3498 CGCCGACCACCTCGTCCTCACCAGCCGCCGCGGCCCCTCGGCCCCCGGCGCGCCCGCGCT 3557

Qy        710 CGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGT 769
              ||| | | | |       | | |||| | || | |  ||  ||     |  || ||| |
Db       3558 CGCCGCCGACCTGGAGGCGCTGGGCGCCCGCGTCACCACCGCCGCCTGCGACGTCGCCGA 3617

Qy        770 CGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCA 829
              |  || | |   ||    ||  ||   |    |    | ||||||    |  |  |
Db       3618 CCGCGCCGCGCTCGCCGCCCTGCTGACCGGCCTGGAGCGCGCGGGCGAGCAGGTCACCGC 3677

Qy        830 CGACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGA 889
              || ||||  ||    ||  | | ||   |    ||   |||  |    ||||   |||||
Db       3678 CGTCGTGCACGCGGCGGGCGCCAACGCCCAGACCCCCGTCGCCGACACCACCCCCGAGGA 3737

Qy        890 TCGCGC 895
              | |||
Db       3738 GCACGC 3743


RESULT 6
US-10-496-351-1
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             9.0%;  Score 81;  DB 7;  Length 14055;
  Best Local Similarity   46.3%;  Pred. No. 1.4e-08;
  Matches  304;  Conservative    0;  Mismatches  350;  Indels    3;  Gaps    1;

Qy        150 CTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTC 209
              | || |  |  | ||     || ||  | | |    ||| || | ||   || ||||
Db      10735 CAGCGGGCCGGGACCGGGCGGCTGCTGGAGCGGCGGTTCCCGGTGTTCCGGGACGCGCTG 10794

Qy        210 CCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGG 269
              | ||  |   ||     | |  |    ||||  |||||  |      |  || |    ||  |
Db      10795 CGCGAGGTGTGCGCCCTGCTGGACCGGCGCGTCGTCGGCGGTCCCGGCGTGCGCGCCGCG 10854

Qy        270 CCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCC 329
              | |  | ||   |  ||  ||| ||||  ||    || ||      |   |      | ||
Db      10855 CTGGACGACCCCGGGCTGCTCGCCGACACCCGCTACGCGCAGGCCGGCCTGTTCGCCGTG 10914

Qy        330 GACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCC 389
              | |  | | | |       | |   | | |||  || || || ||      || ||
Db      10915 CAGGTCGCGCTGGTGCGCCTGCTCGACGCGCTCGGCGTCCGCCCCGACCTGCTCGCCGGG 10974

Qy        390 GACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGC 449
              ||    |  ||||||  |||||     ||  ||   ||||| |    |    |||| |    |
Db      10975 CACTCGGTCGGCGAGATCGCCGTCGCGCACGC---CGCCGGGGTGCTGTCCCTGGAGGAC 11031

Qy        450 CTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTC 509
              |  | |||  | | |     | | ||      |    | | | ||  ||| |  || | |
Db      11032 GCCTCCACCCTCGTGGCCGCGCGCGGCGCCCTCATGCGCGAGCTGCCCCCCGGCGTGATG 11091

Qy        510 TTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTC 569
              || ||| |     | | ||||| | || ||   |||  | |||||| | || |  |||
Db      11092 GTCGCGGTCCGCGCGGGCGAGGACGAGGCGCGCGCCGCGCTGGTCGACGGGGTGGAGCTG 11151
```

```
Qy        570 CTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTC 629
              |  || || ||  |      ||  | || ||| |     ||  | | |   |  ||
Db      11152 GCCGCCGTCAACGGCCCCCGCGCCACCGTGCTCACCGGCGACGAGGCCGCCGTCACCGAG 11211

Qy        630 GGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAG 689
              | |  |  ||  |     |||  |  || |||  ||    |  || ||  || || |
Db      11212 GTCGCCGCCCGCCTCGGGCGCGCCACGCGCCTGCGGGTCGCCCACGCCTTCCACTCGGCC 11271

Qy        690 GGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCC 749
              |   ||| ||  |     | |||| ||      |    || |||||  |  |  |    |
Db      11272 GCCGTCGACGCGGTGCTGCCCGCGTTCGCCGACGTGCTGGCGCGCCTGGACTTCCGCCCG 11331

Qy        750 CGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGG 806
              |     |||  |  ||   |  ||  |||| || ||||  |     |  | |||||
Db      11332 CCGCGCGTCCCGGTGGCCACCACCGCGCCCGGCGCGCTCGACGCGCCCGAGTACTGG 11388


RESULT 7
US-10-496-351-4
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 9975
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-4

  Query Match             8.8%;  Score 79.6;  DB 7;  Length 9975;
  Best Local Similarity   46.8%;  Pred. No. 2.9e-08;
  Matches  386;  Conservative    0;  Mismatches  429;  Indels    9;  Gaps    4;

Qy          5 CCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGG 64
              ||  | || ||   | |  | | |    | |  || |  ||| |  |  |||||||
Db       3645 CCTCGCCGTGGTGCAGGGGTTCCTCGCCGATCCCCGGCACGCCGACTCGGTGCTGCTCGT 3704

Qy         65 GGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGA 124
              ||| | | | ||||   ||  ||||  |  |  |   | ||     || | ||||  | |
Db       3705 GGTCACCAGGGGCG--GCGAGGGCCCGGACGGCGCGACCGCCGCGACGTCCGCCGTGC-A 3761

Qy        125 CGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTC 184
              ||   ||||| ||||   ||| ||   |  || |   |        |||  |  || |
Db       3762 CGGCCTCGTCCGCTCCGCCCAGGCCGAGGAGCCCGGCCGGATCGTCCTGGTGGACACCGA 3821

Qy        185 CCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGT 244
              |  |   || | |  ||| | | || |||    || |   | | |   |  |  | ||
Db       3822 CGGCTCCGCCCCCGACGACCTGATCACCGCGCTCGCCGCCGGACCGGAGCCCCAGCTCGC 3881

Qy        245 CGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCG 304
              |||   ||    | |||   ||  |||| |      |     || |  ||| | |
Db       3882 CGTGCGCGCGGGCCGGGTCACCGCGCCGAGGTTGCGCCGCGCCGCCCCCGAGGGCGCCGC 3941

Qy        305 CGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCT 364
              || |  | | |  || | ||| |  ||  | |||||   | |     |   |||
Db       3942 ACCCGTCCTGGACCCCGAGGGCACCGTCCTGGTCACCGGCGGCACCG---GCGTGCTCGG 3998

Qy        365 CGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGAT 424
              ||    | ||||   | || ||| |  || |   |||| | ||  |  || ||  |
Db       3999 CGCGGCCGTCGCCCGCCACCTCGTCACCGCCCGCGGCGCGCGCCACCTGCTGCTGCTCAG 4058

Qy        425 CGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGA 484
              |  |  |  |||  |  || ||  | || ||   |||    | |  | | |  |
Db       4059 CCGCCGGGGCCCCGACGCGCCCGGCGCGGCCGACCTGGTCGCGGAGCTGACCGGCGCGGG 4118

Qy        485 TCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGAC 544
              || | | |   ||| ||   ||  |   ||||  ||     |    || |   ||
Db       4119 CGCCCGCGTGGTGGTCCGCGCCGTCGACGTCGCCGACCGCGCCGCCCTCGCCGCCGCGCT 4178

Qy        545 CGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAG 604
              |||   ||||    |  |||   | |||| || | ||||||     | ||  |  | ||
Db       4179 CGACGCGGTCCCCGCCGCGCACCCGCTCACCGGCGTCGTGCACCTGGCCGCCGCGCTCGA 4238

Qy        605 CAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGG 664
              | | |   |||     ||  |  | ||| |  ||| |     || ||  | |   |
Db       4239 CGACGGCACGATCACCGCGCTCACCCCCGACCGCGTGCGCCGGGTCCTCGCCCCCAAGGC 4298

Qy        665 GTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGAC---GGATACCTCGCGGTCTATCC 721
              | |   | ||| |||      ||  || |  ||||   |      ||||  | |  |
Db       4299 CACCGCCGCCGCCCACCTGCACGAGCTGACCTCCGACCTGGCGCTGTTCGCCCTGTTCTC 4358

Qy        722 CACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGC 781
              | |       |  ||| ||||| |||| |  ||  |  |  | ||||||  | | || ||
Db       4359 CTCGGCCTCCGGCGTGCTCGGCGCGCCGGGCCAGGGCAACTACGCGGCGGCCAACGCCTA 4418

Qy        782 CGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCG 825
              |  |  | | || | ||   |  |   |||||   || ||  ||
Db       4419 CCTGGACGGGCTGGCCCGCGAGCGCCACGCGGCCGGGCTGCCCG 4462


RESULT 8
US-10-496-351-17
; Sequence 17, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
;   LENGTH: 1026
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-17

  Query Match             8.7%;  Score 78.6;  DB 7;  Length 1026;
  Best Local Similarity   45.2%;  Pred. No. 7.6e-08;
  Matches  365;  Conservative    0;  Mismatches  439;  Indels    3;  Gaps    2;

Qy         11 CGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGAT 70
              ||||  ||   |||          | | || |||| | | | |          ||    | |
Db       219 CGCGCTCGCGCTCTCCAGGGGCCTGCCCGTGCTCCTGGTGGACCCGGACGCCGACCGGCT 278

Qy         71 CGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGT 130
              || ||  |  |  || | || |||    |   | |||   |  |||||  || || ||  |
Db       279 CGACGCGGCCCGCGCCGACGTGCGCGCGCACCTGCGGACCGCCCAGCTGCTCGGCGTCGC 338

Qy        131 CGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCG 190
              || |||  ||| |  || |||| |   | || || |      |  || ||       |
Db       339 CGCCGGGCCGCTCGGCGAGCTGACCACCGCGACCGACACGGGCGGCCCCCGCGAGGTGGT 398

Qy        191 GGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCAT 250
              ||  | ||||| || || |||  | || |  |   |  || | |   | | ||   |
Db       399 CGCCGTGGTCGAGGCGGTCACCGAGGACGCCGAGACCAAGGCCAAGGCGCTCACCGGGGT 458

Qy        251 CGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCT 310
              | ||  |   |  ||   |  || || |   | ||||      |  | |||| | | |
Db       459 CTGCGCCACCGTCCCGCCGG-GCACCCCGCTGGTCTCCAACACCTCCTCCATCCCCATGG 517

Qy        311 CTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCA 370
                  ||||   ||| | | |||    ||||||    |||  | | | | ||    |
```

```
Db        518 GCGAGCTGGCCCCCGCGCTGCCCAGGCC--CGGCGACCTCGTGGGCGCGCACTTCATGAA 575

Qy        371 CCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGC 430

Db        576 CCCGCCGTACCTGATCCCCGCCGTCGAGGTAGCGCGCGGCCCGCTGACCTCGGACGCGGC 635

Qy        431 GTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCA 490

Db        636 GTTCGCCGGGCTCACCGCCCTGCTCGCGCGCCTCGGCCGCGCCCCGGTACAGGTCGGCGA 695

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550

Db        696 CGCGCCGGGCTTCGTCACCAGCCGCCTGCTGCACCCGATGATCAACGACGCGGCGCGCGT 755

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610

Db        756 CGTGGAGTCGGGCACCGCCGACGCCGCCGCCGTCGACGCGCTCATGCGCGACTGCCTCGG 815

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGAC 670

Db        816 CCACCGCGAGGGCCCGCTGCGCACCGCCGACCTGATCGGCATCGACAACCTCGTCGACTC 875

Qy        671 GAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGAC 730

Db        876 GCTGCGCGTGCTGCGCAGGCGCACCGGGGACGAGCGGTGCGAGCCGTGCGACCTGCTGCT 935

Qy        731 GCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCG 790

Db        936 GGCCAAGGTCGACGCGGGGCACCTCGGGCGCAAGAGCGGCCGGGGCTTCTACGACTACGG 995

Qy        791 CCTCGTCCTGCACTGGGGCGCGGTGCG 817

Db        996 CCACCCCGGCCGTGAAGGGACGGGGCG 1022
```

RESULT 9
US-10-496-351-8
; Sequence 8, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2247
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-8

Query Match 8.6%; Score 77.8; DB 7; Length 2247;
Best Local Similarity 44.6%; Pred. No. 9.6e-08;
Matches 307; Conservative 0; Mismatches 382; Indels 0; Gaps 0;

```
Qy        171 CTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGC 230

Db        139 CTGCTCCGCCAGCCCGGCATCACCGGGGCGCAGGCCAGCCCCGCGACCGGCCGGGTGCTG 198

Qy        231 GCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTC 290

Db        199 GTGCTGCACGACCGGGCGCTCGGCGTCGCCGAGGTCGGCGCCGCCCTGCGGGACGCGGTC 258

Qy        291 GTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCG 350

Db        259 GCCGCCGCGGCCAGCGCCCCGCTGGTCGTCGAGGGCAGCACCACGACGAGCCCACCGGCC 318

Qy        351 GTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCC 410

Db        319 ACCGCCGCGAGCACGTCCGGCGGCACGTCCGAGGTCGCGCTCCGGCCCGCGAGCGCCGTC 378

Qy        411 GCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGG 470

Db        379 CTGCGCCGCCCGCTCGCCCTGGTCGGCGTGGTCACCGGCGTCTCGTCCGCCGTCCTGCGC 438

Qy        471 ATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAG 530

Db        439 CACGCCTGGCGCGCCCTCCTGCGGCCCCTGCGCCGCACCGCCGCGGGCACCCGGCTCTCC 498

Qy        531 GAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCG 590

Db        499 GCCCACCCGCTGGTGCGCGCCCTCGGCCGCCGCCGGTCGGCGCTCGCCGCCGCCGCCGCG 558

Qy        591 ACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGC 650

Db        559 CTGACGGTGGCCTGCCAGCTCGCCGAGCTCGGCCTCGGGCTGTTCCTCGGCTGGATCCTG 618

Qy        651 CGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTC 710

Db        619 CTGGTGCTGATCAAGGGCGAGTACGCCCCGCTCACCGCGCTGGGCGTCACGAGCGCCCCC 678

Qy        711 GCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTC 770

Db        679 GCCCAGCTGCTCGCGCTCGCCGCCGGGGCCGCCGCGACCTGCGGGGTCGTCGTCGTGCTC 738

Qy        771 GACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCAC 830

Db        739 TCCGCCGCCTCCGGCGCCCGCTGGCGCGCGCTCGGCCAGGGCGTCCGGCGCGACCAGCGC 798

Qy        831 GACGTGATCGGCAGGGAGGTCCACCGGCT 859

Db        799 GCCCGCCTGCACCGCCACACCCAGCGCCT 827
```

RESULT 10
US-11-205-109-1
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; CURRENT FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 88421
; TYPE: DNA
; ORGANISM: Actinoplanes sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2077)..(3078)
; OTHER INFORMATION: ORF 1; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3118)..(4032)
; OTHER INFORMATION: ORF 2; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4038)..(5048)
; OTHER INFORMATION: ORF 3; positive strandedness
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6665)..(5814)

```
;   OTHER INFORMATION: ORF 4; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7703)..(6693)
;   OTHER INFORMATION: ORF 5; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9464)..(8130)
;   OTHER INFORMATION: ORF 6; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9691)..(10761)
;   OTHER INFORMATION: ORF 7; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (12751)..(10829)
;   OTHER INFORMATION: ORF 8; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13617)..(12802)
;   OTHER INFORMATION: ORF 9; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15203)..(13614)
;   OTHER INFORMATION: ORF 10; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15591)..(15863)
;   OTHER INFORMATION: ORF 11; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15880)..(19035)
;   OTHER INFORMATION: ORF 12; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19032)..(39713)
;   OTHER INFORMATION: ORF 13; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (39713)..(65800)
;   OTHER INFORMATION: ORF 14; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65826)..(66530)
;   OTHER INFORMATION: ORF 15; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (66546)..(67370)
;   OTHER INFORMATION: ORF 16; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (67384)..(70059)
;   OTHER INFORMATION: ORF 17; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70099)..(70662)
;   OTHER INFORMATION: ORF 18; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70659)..(71906)
;   OTHER INFORMATION: ORF 19; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (73439)..(71964)
;   OTHER INFORMATION: ORF 20; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (74216)..(73563)
;   OTHER INFORMATION: ORF 21; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75424)..(74213)
;   OTHER INFORMATION: ORF 22; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75535)..(76464)
;   OTHER INFORMATION: ORF 23; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78110)..(76449)
;   OTHER INFORMATION: ORF 24; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79864)..(78107)
;   OTHER INFORMATION: ORF 25; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81624)..(79861)
;   OTHER INFORMATION: ORF 26; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81909)..(81682)
;   OTHER INFORMATION: ORF 27; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82346)..(82062)
;   OTHER INFORMATION: ORF 28; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82587)..(84446)
;   OTHER INFORMATION: ORF 29; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84481)..(85548)
;   OTHER INFORMATION: ORF 30; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85556)..(86845)
;   OTHER INFORMATION: ORF 31; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87372)..(86803)
;   OTHER INFORMATION: ORF 32; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87494)..(88420)
;   OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

  Query Match             8.6%;  Score 77.4;  DB 14;  Length 88421;
  Best Local Similarity   46.0%;  Pred. No. 5.4e-08;
  Matches  301;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy        188 CCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGT 247
              || | |   |  |  |||||  |||| ||  | |  |  |||||     |  |||  ||
Db      46705 CCTGCCCGCCGACCGGCCGCGGCCCGCCGCGGCCTCGCACCGCGGCCACCTGCACAGCGC 46764

Qy        248 CATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGA 307
                  ||  |  ||        | | |||   |      ||| || ||||  |
Db      46765 GGAGATCCCGGCCGACGTGCACCGGAGCCTGCGCCGGGTCGCCGCCGACCACGGCGCGAC 46824

Qy        308 GCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGA 367
              |||||| |  || | | |||| || || | ||      | ||  | ||  |     |
Db      46825 CGTCTTCATGACCCTGCAGGCCGCCGTGGCGGTCCTGCTGTCGCGGCTCGGCGCGGGCAC 46884

Qy        368 CCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGC 427
              | || || ||  || |  || || || ||   |||  | |   |||| |  |    | |
Db      46885 CGACGTCCCGATCGGCACCGTCGTCGCCGGCCGCGCCGACCGGGCGCTGGAGAACCTGGT 46944

Qy        428 CGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCT 487
              ||  |    |  | ||  |   ||||      | |  ||| | || | ||          ||
Db      46945 CGGCTT---CTTCGTCAACACGCTCGTGCTGCGCACCGACCTGACCGGCGACCCGCGGCT 47001

Qy        488 CCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGA 547
              |  ||  |||  | || |   |  |  | | ||||| |  |||  | | |  ||
Db      47002 GACCGACGTGCTCGGCCAGGTGCGCGAGCTGACCCTGCGGGCGCTGGCCCACCAGGACGT 47061
```

```
Qy        548 TCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAA 607
Db      47062 CCCGTTCGAGAAGCTGGTCGAGGAGCTCACCCCGGCCCGCTCGCTCGCCCGGCACCCCCT 47121

Qy        608 GATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTC 667
Db      47122 GTTCCAGGTCATGGTCACCCTGGACGGCGGCGGGCCGGACGGCGCCGAGCTGCCGGGCCT 47181

Qy        668 GACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCAT 727
Db      47182 GGCGATGTCGGTCGTGCCGACCGGCGCCGTTCCGGCCAAGTTCGACCTCGACCTCACGTT 47241

Qy        728 GACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCT 787
Db      47242 CACCGAGACCTTCGACGCCGCGGGGGAGCCGGCCGGCCTGCGCGTCGACCTCATCGCGGC 47301

Qy        788 CCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGATCGGC 842
Db      47302 GGCCGACCTCTTCGACGCGGGCACGGCCGCCCGGCTCGCCGGCTACCTGAGCCGC 47356
```

RESULT 11
US-10-496-351-12
; Sequence 12, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 3354
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-12

Query Match 8.6%; Score 77.2; DB 7; Length 3354;
Best Local Similarity 45.5%; Pred. No. 1.2e-07;
Matches 347; Conservative 0; Mismatches 413; Indels 2; Gaps 2;

```
Qy         53 CGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCG 112
Db       1422 CGCTGTGCTGGACGTGCTGCCCGCCGCCCGGCGCGCCGAGCTGCACCTGCGCGCCGCGCG 1481

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172
Db       1482 GACGTTGCACGAGCGCGGTGAGTCGCCCGCCGTCGTCGCCGACCACCTCGTCACCGCCGG 1541

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232
Db       1542 GCTGCCCGCGCCCGCCTGGGCCGTCGACGTGCTCGTCGCCGCCGCCGACCACGCGCTCGC 1601

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292
Db       1602 CGCCGACGACGTGCCCACCGCCGTCGCACGGCTCGCCGCCGCCCACCGCCGGGACGACGC 1661

Qy        293 CGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGT 352
Db       1662 GGACCAACCGCGGGCCGTCCCCCCGCGCGCAGTCCCGGACCCGCCCGTCGCCGCCCACCG 1721

Qy        353 GTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGC 412
Db       1722 GTGCGGGTGCGCGGACCGGACGGCCGCGCCCGCCGAGCGCCGC-ACGTCGCCCGAGGCCG 1780

Qy        413 GCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGAT 472
Db       1781 CCGCCGTCGCCGGGCCCGCGCTCACCGCCCGGCTCGCCCGCGCGCGCTGGCAGCTCAACC 1840

Qy        473 CCGCTGGGACGATCT-CCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGG 531
Db       1841 CGGCTGCCGCGCTGTGCCACCTCGACCCGCTCGCCGACGCGCTCGCCGCCGACCCCGACG 1900

Qy        532 AGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGA 591
Db       1901 CGCTCGCCCCGCGGGACGCCGCCGCGCTGGTGCGCCAGCTGGTCTGGCACGGGCGCACCA 1960

Qy        592 CCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCC 651
Db       1961 CCGCCGCCGGGCGCGTGCTCGCCGCGCTGCGCGCCCGCGAGCACGCCGCCGAGCTGCCCG 2020

Qy        652 GCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCG 711
Db       2021 ACCTGGAGACCTGGCTCTCCCTCGCCTGCCCGCCGCTCACCGCCTCCCGCCGCCCCACAA 2080

Qy        712 CGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCG 771
Db       2081 CCCTCCACGCCTCCGCCCCGCACACCTCCGCCCCGCGCACCTCGGCCCACCCCACCCCCC 2140

Qy        772 ACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGG 813
Db       2141 TCGCCCCACCCCCGAGCCGCGTCGACCCGTGGTTGCGCGAGG 2182
```

RESULT 12
US-10-858-730-38
; Sequence 38, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 2736
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-858-730-38

Query Match 8.5%; Score 76.2; DB 8; Length 2736;
Best Local Similarity 45.4%; Pred. No. 2e-07;
Matches 374; Conservative 0; Mismatches 428; Indels 21; Gaps 2;

```
Qy         27 CACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCG 86
Db        268 CACCGCGGCCGCGAGCTGGGCGCCAAGCGCGCCGCCGAGGGCGGACTGCTCGCCCGTACG 327

Qy         87 GCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCAC 146
Db        328 GCCGACCGGCTGAAGGACGCCGACCCCGAGCACCTGCGCGAGACGGTCCGCAACCTCAAC 387

Qy        147 GCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCG 206
```

```
Db        388 GTGCGCCCCGTGTTCACCGCGCACCCCACCGAGGCCGCCCGCCGCTCCGTCCTCAACAAG 447

Qy        207 CTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCC 266

Db        448 CTGCGCCGCATCGCCGCCCTCCTGGACACCCCGGTCAACGAGTCGGACCGGCGCCGCCTG 507

Qy        267 TGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGA---------GCTCTTCAAG 317

Db        508 GACACCCGCCTCGCCGAGAACATCGACCTCGTCTGGCAGACCGACGAGCTGCGCGTCGTG 567

Qy        318 CGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCG 377

Db        568 CGCCCCGAGCCCGCCGACGAGGCCCGCAACGCCATCTACTACCTCGACGAGCTGCACCTG 627

Qy        378 GCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCC 437

Db        628 GGCGCCGTCGGCGACGTCCTCGAAGACCTCACCGCCGAGCTGGAGCGGGCCGGCGTCAAG 687

Qy        438 ACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCACCTCGCC 497

Db        688 CTCCCCGACGACACCCGCCCCCTCACCTTCGGCACCTGGATCGGCGGCGACCGCGACGGC 747

Qy        498 CTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGG------------AGGCGGAGCCGACC 545

Db        748 AACCCCAACGTCACCCCCCAGGTGACCTGGGACGTCCTCATCCTCCAGCACGAGCACGGC 807

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605

Db        808 ATCAACGACGCCCTGGAGATGATCGACGAGCTGCGCGGCTTCCTCTCCAACTCCATCCGG 867

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665

Db        868 TACGCCGGTGCGACCGAGGAACTGCTCGCCTCGCTCCAGGCCGACCTGGAACGCCTCCCC 927

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725

Db        928 GAGATCAGCCCCCGCTACAAGCGCCTCAACGCCGAGGAGCCCTACCGGCTCAAGGCCACC 987

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785

Db        988 TGCATCCGCCAGAAGCTGGAGAACACCAAGCAGCGCCTCGCCAAGGGCACCCCCCACGAG 1047

Qy        786 CTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATC 828

Db       1048 GACGGCCGCGACTACCTCGGCACCGCCCAGCTCATCGACGACC 1090


RESULT 13
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Staffa, Alfredo
;  TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
;  FILE REFERENCE: 3002-2US
;  CURRENT APPLICATION NUMBER: US/11/205,109
;  CURRENT FILING DATE:  2005-08-17
;  PRIOR APPLICATION NUMBER: US/09/976,059
;  PRIOR FILING DATE: 2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 88421
;   TYPE: DNA
;   ORGANISM: Actinoplanes sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2077)..(3078)
;   OTHER INFORMATION: ORF 1; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3118)..(4032)
;   OTHER INFORMATION: ORF 2; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4038)..(5048)
;   OTHER INFORMATION: ORF 3; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6665)..(5814)
;   OTHER INFORMATION: ORF 4; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7703)..(6693)
;   OTHER INFORMATION: ORF 5; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9464)..(8130)
;   OTHER INFORMATION: ORF 6; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9691)..(10761)
;   OTHER INFORMATION: ORF 7; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (12751)..(10829)
;   OTHER INFORMATION: ORF 8; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13617)..(12802)
;   OTHER INFORMATION: ORF 9; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15203)..(13614)
;   OTHER INFORMATION: ORF 10; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15591)..(15863)
;   OTHER INFORMATION: ORF 11; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15880)..(19035)
;   OTHER INFORMATION: ORF 12; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19032)..(39713)
;   OTHER INFORMATION: ORF 13; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (39713)..(65800)
;   OTHER INFORMATION: ORF 14; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65826)..(66530)
;   OTHER INFORMATION: ORF 15; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (66546)..(67370)
;   OTHER INFORMATION: ORF 16; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (67384)..(70059)
;   OTHER INFORMATION: ORF 17; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70099)..(70662)
;   OTHER INFORMATION: ORF 18; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70659)..(71906)
;   OTHER INFORMATION: ORF 19; positive strandedness
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (73439)..(71964)
;   OTHER INFORMATION: ORF 20; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (74216)..(73563)
;   OTHER INFORMATION: ORF 21; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75424)..(74213)
;   OTHER INFORMATION: ORF 22; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75535)..(76464)
;   OTHER INFORMATION: ORF 23; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78110)..(76449)
;   OTHER INFORMATION: ORF 24; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79864)..(78107)
;   OTHER INFORMATION: ORF 25; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81624)..(79861)
;   OTHER INFORMATION: ORF 26; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81909)..(81682)
;   OTHER INFORMATION: ORF 27; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82346)..(82062)
;   OTHER INFORMATION: ORF 28; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82587)..(84446)
;   OTHER INFORMATION: ORF 29; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84481)..(85548)
;   OTHER INFORMATION: ORF 30; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85556)..(86845)
;   OTHER INFORMATION: ORF 31; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87372)..(86803)
;   OTHER INFORMATION: ORF 32; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87494)..(88420)
;   OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

  Query Match             8.4%;  Score 75.8;  DB 14;  Length 88421;
  Best Local Similarity   47.1%;  Pred. No. 1.2e-07;
  Matches  267;  Conservative    0;  Mismatches  297;  Indels    3;  Gaps    1;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              || |||||        |||   ||    ||| ||| ||  || |  | |      |
Db      80719 CGTCCGGGCGTTCACCGAGCGCTTCGCCCCCTTCGGCCTGCACACCCACGCGGTGAACCC 80660

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              | |     |   | |||     | ||  |   |  | || ||||    ||||
Db      80659 CTGCTACGGCATGGCCGAGTTCACCGCGTACGTGTCGACGAAGGTCTTCGAGGCGCCGGC 80600

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              | |    |  || |   |  | |||||||  |  | |  | |       ||| | |||
Db      80599 GGTCTTTCTTCCCGCCGACCCTCGCGCGCTGGAGGACGCCGCGTCGCCGGCCCTGCGCCC 80540

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              | ||  ||  ||       |  | | | || || | |||||| | | ||   |
Db      80539 GGCCGACCCCGCCGCGGCCCGGGAGATACCGGGTGTCGGCCGGGTGCCCGACTTCGAGGT 80480

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTC-- 299
              || || ||| ||  |  | ||   |||||  | |  ||  |  |||| |||  ||
Db      80479 GCTCATCGTCGACCCGGACGGGCTACGGCCGCTGCCCGAGGGCCGGGTCGGCGAGATCTG 80420

Qy        300 -CAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTC 358
               | ||||| || |  |  | || | | |  |    |    || ||             |
Db      80419 GCTGCGCGGGCCCGGCGCGGGCGCCGGCTACTGGGGCAGGACCGAGCTCAACCCCGGCAT 80360

Qy        359 CCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCG 418
              | ||  || |   | ||||| ||||| |   |||||||   | | | |  | |      |
Db      80359 CTTCGACGCCAGGCCCGCGGGCGACGGCCAGGACGGCGGCTGGGTGCGAACGGGTGACCT 80300

Qy        419 GGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTG 478
              ||      |   |  ||  || |  |||   || |   || | | || |
Db      80299 GGGTGCGCTGACCGGAGGCGAGCTGTTCCTCACCGGACGCCTCAAGGAGCTGCTCATCGT 80240

Qy        479 GGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGA 538
              | |||  | | |||| ||||  | ||| |||   | ||  || |||   | | | |  |
Db      80239 GCACGGCCGCAACCTGGCCCCGCACGACCTCGAGCGGGAGGCCCGGGCCGCGCACGACGC 80180

Qy        539 GCCGACCGATCGGGTCGAGGTGGCGAT 565
              |  |  | | | | ||| || |||| |
Db      80179 GGTGGACCACCAGATCGGGGCGGCGTT 80153


RESULT 14
US-11-075-185-34
; Sequence 34, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34
;   LENGTH: 11070
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-34

  Query Match             8.3%;  Score 74.8;  DB 14;  Length 11070;
  Best Local Similarity   45.0%;  Pred. No. 2.9e-07;
  Matches  375;  Conservative    0;  Mismatches  447;  Indels   12;  Gaps    2;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              || |  | |  | |||  || | | ||   ||  |  || | ||  | ||| |||  |
Db       6249 CGACGCCGCCTGCGCCGAGCTCGACCGCCACCTCGACCGGCCCCTCATGAGCGTCCTCTT 6308

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              ||  |    ||||  |||  | |||||   ||    |    |     ||     |||||
Db       6309 CGCCGACGCCGGCTCCGAGGCCGCGGCGCTGCTCGACCAGACAGCCTGGGCTCAGCCCGC 6368

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
              |       |   |||    | || |   |  | | |      |||    ||| | || |
Db       6369 CCTGTTCGCTCTCGAGGTGGCCCTCTACCGCCAGTGGGATGCCTGGGGCCTGCGCCCCGA 6428

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
              |   || |  | | || |       | || || || ||   ||  ||   ||| | |
Db       6429 GCTGCTGCTCGGCCACAGCATCGGCGAGCTCGCCGCCGCCCACATCGCCGGCGTGCTCGA 6488
```

```
Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
Db       6489 CCTCGCCGACGCCTCCGCCCTGGTCGCCGCCCGCGGGCGGCTCATGCAGGCCCTCCCCCT 6548

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
Db       6549 CGGCGGCGCCAT---------GGCCTCCGTCGAGGCCACCGAGGACGAGCTACGCCCCTT 6599

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
Db       6600 GCTCGACCAGCACACAGGACGCCTCTCGCTCGCCGCCCTCAACACCCCACGCCAGTCGGT 6659

Qy        422 GAT---CGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTG 478
Db       6660 CGTCAGCGGCGACGAGCCCGCCGTCGACCAAGTCTGCGCCCACTTCACCGCCCTCGGCCG 6719

Qy        479 GGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGA 538
Db       6720 ACGCGCCAAGCGGCTCGTCGTCAGCCACGCCTTCCACTCGGCGCACATGGAGCCCATGCT 6779

Qy        539 GCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACAT 598
Db       6780 CGACGCCTTCGCCCGCGTCGCTCGCGGCCTGACCTTCCACCCGCCCCGGCTGCCCATCGT 6839

Qy        599 GATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCT 658
Db       6840 CAGCAGCGTCACCGGCGCACGCGCCTCCGCCGACGAGCTCACCTCGCCCGACTACTGGGT 6899

Qy        659 CCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTA 718
Db       6900 CCGCCAGGTGCGCGAGCCCGTCCGCTTCGTCGACGGCATGCGCGCACTGCACGCCGCCGG 6959

Qy        719 TCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
Db       6960 CGCGGCCACCTTCGTCGAGTGCGGGCCGCACGGCGTGCTCAGCGCCGCCGGCGCAGAGTG 7019

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGA 832
Db       7020 CCTCGCTCCCGACGGCGCTCGCGACGCCGGCTTCGTCCCCAGCCTCCGCAACGA 7073


RESULT 15
US-11-075-185-1
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
;  APPLICANT: REEVES, CHRISTOPHER  D
;  APPLICANT:  JULIEN, BRYAN
;  APPLICANT:  REID, RALPH
;  TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
;  FILE REFERENCE: 010099.03
;  CURRENT APPLICATION NUMBER: US/11/075,185
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/551,103
;  PRIOR FILING DATE: 2004-03-08
;  PRIOR APPLICATION NUMBER: US 60/568,290
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 78869
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-1

  Query Match             8.3%;  Score 74.8;  DB 14;  Length 78869;
  Best Local Similarity   45.0%;  Pred. No. 1.9e-07;
  Matches  375;  Conservative    0;  Mismatches  447;  Indels   12;  Gaps    2;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
Db      10450 CGACGCCGCCTGCGCCGAGCTCGACCGCCACCTCGACCGGCCCCTCATGAGCGTCCTCTT 10509

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
Db      10510 CGCCGACGCCGGCTCCGAGGCCGCGGCGCTGCTCGACCAGACAGCCTGGGCTCAGCCCGC 10569

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
Db      10570 CCTGTTCGCTCTCGAGGTGGCCCTCTACCGCCAGTGGGATGCCTGGGGCCTGCGCCCCGA 10629

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
Db      10630 GCTGCTGCTCGGCCACAGCATCGGCGAGCTCGCCGCCGCCCACATCGCCGGCGTGCTCGA 10689

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
Db      10690 CCTCGCCGACGCCTCCGCCCTGGTCGCCGCCCGCGGGCGGCTCATGCAGGCCCTCCCCCT 10749

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
Db      10750 CGGCGGCGCCAT---------GGCCTCCGTCGAGGCCACCGAGGACGAGCTACGCCCCTT 10800

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
Db      10801 GCTCGACCAGCACACAGGACGCCTCTCGCTCGCCGCCCTCAACACCCCACGCCAGTCGGT 10860

Qy        422 GAT---CGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTG 478
Db      10861 CGTCAGCGGCGACGAGCCCGCCGTCGACCAAGTCTGCGCCCACTTCACCGCCCTCGGCCG 10920

Qy        479 GGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGA 538
Db      10921 ACGCGCCAAGCGGCTCGTCGTCAGCCACGCCTTCCACTCGGCGCACATGGAGCCCATGCT 10980

Qy        539 GCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACAT 598
Db      10981 CGACGCCTTCGCCCGCGTCGCTCGCGGCCTGACCTTCCACCCGCCCCGGCTGCCCATCGT 11040

Qy        599 GATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCT 658
Db      11041 CAGCAGCGTCACCGGCGCACGCGCCTCCGCCGACGAGCTCACCTCGCCCGACTACTGGGT 11100

Qy        659 CCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTA 718
Db      11101 CCGCCAGGTGCGCGAGCCCGTCCGCTTCGTCGACGGCATGCGCGCACTGCACGCCGCCGG 11160

Qy        719 TCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCC 778
Db      11161 CGCGGCCACCTTCGTCGAGTGCGGGCCGCACGGCGTGCTCAGCGCCGCCGGCGCAGAGTG 11220

Qy        779 CGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGA 832
Db      11221 CCTCGCTCCCGACGGCGCTCGCGACGCCGGCTTCGTCCCCAGCCTCCGCAACGA 11274


Search completed: April 27, 2006, 15:53:01
Job time : 566.78 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 4465.48 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_75000_75900
Perfect score: 901
Sequence: 1 acggccgggacgcggacgac.....gaggaggatcgcgccgcgga 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	97.6	10.8	1406	10	CG756569	CG756569 P051-4-B0
c	2	95.6	10.6	1853	10	AG382854	AG382854 Mus muscu
	3	94.8	10.5	2313	7	CK423624	CK423624 AUF_IpSpn
c	4	94.4	10.5	1569	10	AG341503	AG341503 Mus muscu
c	5	94.2	10.5	1436	10	AG435568	AG435568 Mus muscu
	6	92.2	10.2	1844	10	AG304707	AG304707 Mus muscu
	7	90.6	10.1	1785	9	CC219595	CC219595 CH261-62D
	8	90.2	10.0	1798	10	AG171124	AG171124 Pan trogl
	9	90	10.0	1674	10	CL078342	CL078342 CH216-149
	10	89.8	10.0	1006	10	AG058789	AG058789 Pan trogl
c	11	89.8	10.0	1319	10	CL498921	CL498921 SAIL_661_
	12	89.6	9.9	1474	10	CL500927	CL500927 SAIL_68_H
c	13	89.4	9.9	1116	2	BG810038	BG810038 mgct002xi
c	14	89.4	9.9	1116	9	BZ569478	BZ569478 pacs2-164
c	15	89.4	9.9	1821	10	CL090560	CL090560 ISB1-17N1
c	16	88.6	9.8	1878	10	AG393083	AG393083 Mus muscu
	17	88.2	9.8	1190	5	BQ673485	BQ673485 AGENCOURT
c	18	87.6	9.7	1309	10	AG077201	AG077201 Pan trogl
c	19	87.6	9.7	1542	10	AG032943	AG032943 Pan trogl
c	20	87.6	9.7	1628	10	CG757066	CG757066 P052-2-A0
	21	87.2	9.7	1780	10	AG448243	AG448243 Mus muscu
c	22	87	9.7	1137	2	BG809979	BG809979 mgct002xd
c	23	87	9.7	1506	10	AG442117	AG442117 Mus muscu
	24	86.8	9.6	1538	10	AG030607	AG030607 Pan trogl
c	25	86.6	9.6	1692	10	AG396765	AG396765 Mus muscu
	26	86.4	9.6	1232	10	AG072425	AG072425 Pan trogl
c	27	86.2	9.6	1080	9	AQ893056	AQ893056 HS_4832_A
	28	86.2	9.6	1218	10	CL961835	CL961835 OsIFCC037
c	29	86.2	9.6	2243	10	AG381986	AG381986 Mus muscu
	30	86	9.5	2332	10	AG363333	AG363333 Mus muscu
	31	85.8	9.5	1393	10	CL073885	CL073885 CH216-130
c	32	85.8	9.5	1438	9	BZ557931	BZ557931 pacs1-60_
c	33	85.4	9.5	949	10	AG171092	AG171092 Pan trogl
	34	85.2	9.5	764	10	AG060189	AG060189 Pan trogl
	35	85.2	9.5	983	8	CV769389	CV769389 FGAS06378
	36	85.2	9.5	1165	10	AG030649	AG030649 Pan trogl
c	37	85.2	9.5	1956	10	CG754548	CG754548 P050-1-A1
	38	85	9.4	957	10	AG062993	AG062993 Pan trogl
	39	85	9.4	1048	7	CK408336	CK408336 AUF_IfLvr
	40	85	9.4	1885	10	AG363563	AG363563 Mus muscu
c	41	84.8	9.4	2332	10	AG363333	AG363333 Mus muscu
	42	84.6	9.4	989	10	CL041117	CL041117 CH216-51N
c	43	84.6	9.4	1822	10	AG435170	AG435170 Mus muscu
	44	84.4	9.4	1131	10	AG042920	AG042920 Pan trogl
	45	84.4	9.4	1392	10	AG390642	AG390642 Mus muscu

ALIGNMENTS

RESULT 1
CG756569
LOCUS CG756569 1406 bp DNA linear GSS 24-OCT-2003
DEFINITION P051-4-B06.zb Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG756569
VERSION CG756569.1 GI:37984265
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1406)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1406
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 10.8%; Score 97.6; DB 10; Length 1406;
Best Local Similarity 46.2%; Pred. No. 4.2e-08;
Matches 392; Conservative 0; Mismatches 439; Indels 17; Gaps 2;

Qy 53 CGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCG 112

```
Db        487 CCGCCCGNCCGGCCGGGGCCGCGCCCCGCCCGGGGCCGCGCGCGGCCGGCCCCGCGGGGG 546

Qy        113 CCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCT 172

Db        547 CCCCCCGGCCGGCGGCCGCGGGGGCGCCCCCGCGCCCCCCCCGCCCCCGCCCCCGCCCCG 606

Qy        173 CCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGC 232

Db        607 CCGCCGGCCGGCCCCCCGCCCGGCCCCGGCCCCGCGCCCCCGCCCGGCGCCGCCCCGCCC 666

Qy        233 GATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGT 292

Db        667 CCCCCCGCCCCGGCGCGGCCCCCGGCGCGCCCGCGGCCCGCGCCCGCCCGCGCCCCGCGC 726

Qy        293 CGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGT 352

Db        727 CCCCCCCCCCCCCGGGGCCGCCGCCCCCCCGCGCCCCGGCCCCCCGGCCCCCCGCGCGG- 785

Qy        353 GTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGC 412

Db        786 -----------CCCGCGCCGCGGGGCGGCCGGGGCCCGGCGCCCGGCCCCGGCCCGCCG 833

Qy        413 GCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGAT 472

Db        834 GGGGCCCCCGCCGCCCCCCCCCCCCGGCCCCCGCGCGCGCGCCCCGCCCGCCGCGCCCGG 893

Qy        473 CCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGA 532

Db        894 GCCCCCGCCCGCGCCCCCCGGCGGCCGCCCCGCCCGGCCCCCCGCCCCCGCCCGCGCGGC 953

Qy        533 GGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGAC 592

Db        954 GCGGCGNNCGGCCGCCGCGCGCGGCGCCGCCGCGCCGCCCGCCCCCGCCCGCCCCCNCCCC 1013

Qy        593 CCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCG 652

Db       1014 CGGCGGCGCCCGNCCGCGCGGCGGCCCCCGCCGCGGCCCCCGGCCCC----CCCCGGCCC 1069

Qy        653 CGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGC 712

Db       1070 CGCGCCCCGGCGGNCGCCCCCCCCGCGCCGGCGGCGCCCGCCCCCGGGCGCGCCGCCGCG 1129

Qy        713 GGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGA 772

Db       1130 CGCCGCCCCCGCGCGCCCGCGCGCGCGCCCCGCCGCGCCCGCCCGCCCGCGGGCCGCGGG 1189

Qy        773 CGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGA 832

Db       1190 CCCCCCCCCCGCCGCCCCGCGCGCCGNCCGCGCCCCGGGCGCCCCGCCCGGCCCCGNGCG 1249

Qy        833 CGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCACCGAGGAGGATCG 892

Db       1250 CGCGGCCGCCCGCGCGCCGCCCCCGCCCCNCCGCCCGCGGGCNGCGCGCGCGCCGCCG 1309

Qy        893 CGCCGCGG 900

Db       1310 CCCCGCGG 1317


RESULT 2
AG382854/c
LOCUS       AG382854               1853 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-193K01.TJ, genomic survey
            sequence.
ACCESSION   AG382854
VERSION     AG382854.1  GI:47994059
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1853)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1853
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-193K01.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             10.6%;  Score 95.6;  DB 10;  Length 1853;
  Best Local Similarity   47.5%;  Pred. No. 9.5e-08;
  Matches  404;  Conservative    0;  Mismatches  439;  Indels    8;  Gaps    4;

Qy         17 CGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGGGTGATCGGCGG 76

Db       1638 CAACAGCACACACCGACACAGCCAGCGCGTGCTGCGCGTGCTGCCCCGCCGCGGCGTCAC 1579

Qy         77 CGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGG 136

Db       1578 CCGNCTCGCGCGGCGGCGGCCGCGTGCCCGGCGCCGCCGCGCGCGCCCTCCGCCTGCGGC 1519

Qy        137 CTCGCGCCACGCGCTGCCGCGCTCGC--CCTATTACCTCCTCCACCCGTCCCTCCGGGCG 194

Db       1518 CGCGCGCCGCCCGCTCCCGCGCCCGCCGCCCGCGGCCGCCGGCCGCCGGGCCCCCGNGGG 1459

Qy        195 CTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGC 254

Db       1458 CGCCGCCGGCCGCGCGCTCGGCGGCGCCGCCCGCCCCGCGCGCGCCCCGCCGCCGGCG-- 1401

Qy        255 CACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTC 314

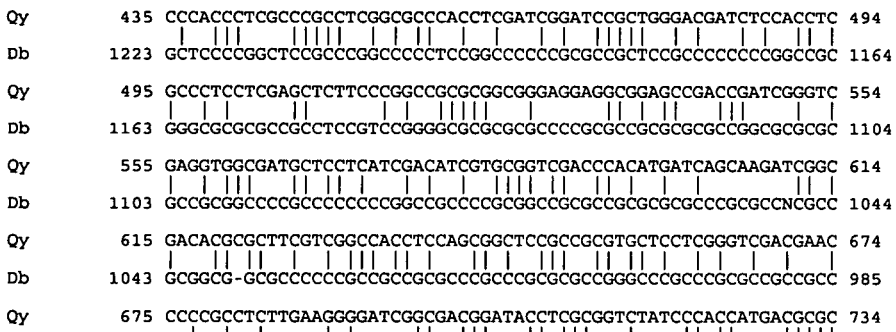
Db       1400 --CGGTCCGCGCGGCGCCCGGCGCCGGCTCCGCCCGCGGCGCGCGCCGCCGCCGCCCGCC 1343

Qy        315 AAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTC 374

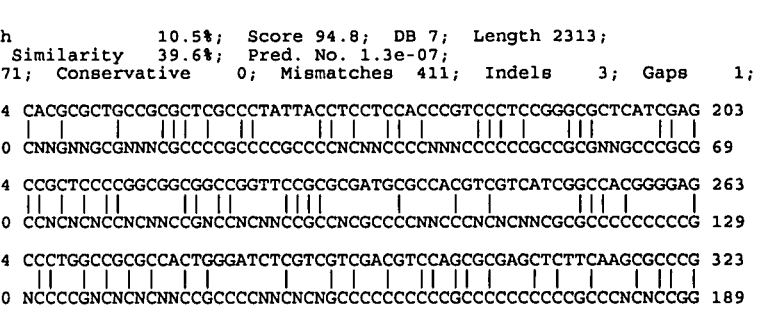
Db       1342 CCCCGCCCGGGCGTCCGGGGCGCGCNCGCCGCCCCGGGCGGCCCGGCGCCGCCCCGGCGC 1283

Qy        375 GCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCA 434

Db       1282 GCGCCCGCCGCCCCCGCGCGCCTCCGCGGCCGC-CCGCGCCGCCCCCCGCCCGCGGCCGCC 1224
```



RESULT 3
CK423624
LOCUS CK423624 2313 bp mRNA linear EST 06-JAN-2004
DEFINITION AUF_IpSpn_68_m14 Spleen cDNA library Ictalurus punctatus cDNA 5' similar to nuclear antigen EBNA1 [Human herpesvirus 4], mRNA sequence.
ACCESSION CK423624
VERSION CK423624.1 GI:40663411
KEYWORDS EST.
SOURCE Ictalurus punctatus (channel catfish)
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes; Ictaluridae; Ictalurus.
REFERENCE 1 (bases 1 to 2313)
AUTHORS Liu,Z., Li,P., Liu,L., He,C., Kucuktas,H., Feng,J., Chen,L., Peatman,E., Baoprasertkul,P., Simmons,M., Muir,W., Grizzle,J., Dunham,R. and Brady,Y.
TITLE 30,000 new catfish ESTs: new resources for functional analysis of genes involved in aquaculture performance traits
JOURNAL Unpublished (2004)
COMMENT Contact: Liu ZJ
The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel: 334 844 4054
Fax: 334 844 9208
Email: zliu@acesag.auburn.edu
Seq primer: T7.
FEATURES Location/Qualifiers
source 1. .2313
/organism="Ictalurus punctatus"
/mol_type="mRNA"
/db_xref="taxon:7998"
/clone_lib="Spleen cDNA library"
/note="Organ: Spleen; Vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN



RESULT 4
AG341503/c
LOCUS AG341503 1569 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-134O11.T7, genomic survey sequence.
ACCESSION AG341503
VERSION AG341503.1 GI:47914813
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end

```
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1569)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1569
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-134O11.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             10.5%;  Score 94.4;  DB 10;  Length 1569;
  Best Local Similarity   46.8%;  Pred. No. 1.6e-07;
  Matches  372;  Conservative    0;  Mismatches  417;  Indels    6;  Gaps    4;

Qy         54 GGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGC 113
Db       1459 GCCCCGCGNNGNCGCGCCGCCGCGCGNGCNNCGNCCCCGCGCGCCGCGCCCCGCGGCGGC 1400

Qy        114 CAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTC 173
Db       1399 GCGCCCGCGGCGCGGCCGGCGCGCCCCCGCGCCGCGCNCGCGGCGCGGCNGNCCNGNCGC 1340

Qy        174 CTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCG 233
Db       1339 CGCCGCGCCGCNGCGCCCCCCGCGCGCGGCCCGCCGCCCGGNCCGGGCGCGCNGCGNCCG 1280

Qy        234 ATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTC 293
Db       1279 CGNGGCCCCGNCGGCGCGCCGCGCGGCGCGCGCGGCGCCGGGCGCGCCCGCCGNCGCGCC 1220

Qy        294 GACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTG 353
Db       1219 GNGCGCCCGGCCGGCGNCGCGNCGCCGGCNCCGGCCGCGGCCGCCCCGCCCGCCCGGCGC 1160

Qy        354 TTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCG 413
Db       1159 CGGGGCCGCGGCCCCCGCCCCGGNCCGGCGGGCGCCGCCGGGCCCGGCCCCCGCCCCGCC 1100

Qy        414 CGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATC 473
Db       1099 GCNCGCGCGCGCGCCGGCGCCCCCGCCCCCGCCGCCGCGCGGCCGCGCGGCGCCGCCGCC 1040

Qy        474 CGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAG 533
Db       1039 CCGCGCCCCGCGCGCCCGGCCGCCCCCCGCGCGCGC-CCCCGGCCCCGCCGCGCGGCCGC 981

Qy        534 GCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACC 593
```

```
Db        980 CCGCCGCCGCGCCCGC-CCCCCCGGCGCCCCCGCGCCGCCGGGNGGCCGCCGCCCCGCGG 922

Qy        594 CACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGC 653
Db        921 CGCCCCGCGGGCCCGCGCGCGGCCCCCCCGCGGNCCCCGGCGCCGCCCGGCGC-CCGCCCC 863

Qy        654 GTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCG 713
Db        862 CGNCCCCGCGGCCCGCCGGGCCCCCGCCGCGCCCGCCGCCGGGCGCGCCGCGCCCCCGCG 803

Qy        714 GTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGAC 773
Db        802 CGGGNGGNCCCC---CCGCGCGCGCCCGGGGCCGCCCGCGCGCCCGCGGGCGCCCGCGGC 746

Qy        774 GACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGAC 833
Db        745 GGCCGCCCCGCNCGCCCGCCCGGGCCGCCGGGGGGCGCCGGGCGGGGCGCCGCCCGCGCC 686

Qy        834 GTGATCGGCAGGGAG 848
Db        685 GCCGGCGGCGCGGCG 671
```

```
RESULT 5
AG435568/c
LOCUS       AG435568               1436 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-313L12.T7, genomic survey
            sequence.
ACCESSION   AG435568
VERSION     AG435568.1  GI:48078631
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1436)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1436
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
```

```
                     /clone="MSMg01-313L12.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             10.5%;  Score 94.2;  DB 10;  Length 1436;
  Best Local Similarity   47.1%;  Pred. No. 1.8e-07;
  Matches  343;  Conservative    0;  Mismatches  384;  Indels    2;  Gaps    2;

Qy         72 GGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTC 131
Db       1414 GGCGGCGCCGCCCGCCGCCCGCCGCGCGCGCCGCGGCNCCNGCCGCCGGCGCCCGCCCGC 1355

Qy        132 GTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCGG 191
Db       1354 CGCCCCGCCGCCGCCGCCCCGCGCGCGGCCGCCCCCGCGCGCCGCGCCCCCCCCGCGCC 1295

Qy        192 GCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCATC 251
Db       1294 GCCCGCCCCGGCGCGCGCCCGCCCGGCGCGCGCCCCCGCCGGCGGCCCCGCCCCGGGCGC 1235

Qy        252 GGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTC 311
Db       1234 CCCGCCCCGCCCCCCCCGNCCGCCGCGCCNGCCGCCCCCCCCGCCCCCCGCCGNCCGCCC 1175

Qy        312 TTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCAC 371
Db       1174 GCCGGCGGGGCCGCCCCCCCCCGCCGCCGCGCCCCGGCGCGCGCCCCCGCGGCACGCCGC 1115

Qy        372 CTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCG 431
Db       1114 CGCCCCCGCGGCCCGCGCGCGCCCGCGGCGGGCCGCCCCGCGCGCCCGCCCCGCGCCGCC 1055

Qy        432 TCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGG-GACGATCTCCA 490
Db       1054 GCGCCCCCCCCCCGCCCGCCCGCCCCGGCGCCCGCCCGGCCCCGCGGGCGGCGCCCGCGC 995

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550
Db        994 GCGCGCGCCGCCCGCCCCCCCGCCCGCCGCGCGGCGCCGCCGCCCCCGCGCCCCGGGCCG 935

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610
Db        934 GCGGCCGCCGGCCGCCCGCGCGCGGGCCGCCGGCCNCGCGCCCCGCGGCGCCCGGGCGGG 875

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTC-CGCCGCGTGCTCCTCGGGTCGA 669
Db        874 CCGCGCCGCCCCCGCCGCCCGCCGCCCCCGGCGCCGCGCGCCGCCCCCCCCCGGGCCCGC 815

Qy        670 CGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGA 729
Db        814 CGCCCGCGCCCCCCGCCCGCGCCCGCGGCGCCGCCGGCCGGGCCCGCCCCCGCGCCGGGC 755

Qy        730 CGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCC 789
Db        754 GCCCCGCGCCGCCGCCGCCCCGCGCGCCCGCCCCCCCCGCCGCCGCCCGCGGCCCGCCCC 695

Qy        790 GCCTCGTCC 798
Db        694 CGCCGGCCC 686


RESULT 6
AG304707
LOCUS       AG304707               1844 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-085A01.TJ, genomic survey
            sequence.
ACCESSION   AG304707
VERSION     AG304707.1  GI:47877661
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1844)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1844
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-085A01.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             10.2%;  Score 92.2;  DB 10;  Length 1844;
  Best Local Similarity   46.1%;  Pred. No. 4e-07;
  Matches  347;  Conservative    0;  Mismatches  403;  Indels    3;  Gaps    1;

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
Db        503 CGCGTGGCCCGGCCGCGCGCCGCCCGCCCCGCGCCCCCCCCGCCCGCCCCGGGCCTGGCC 562

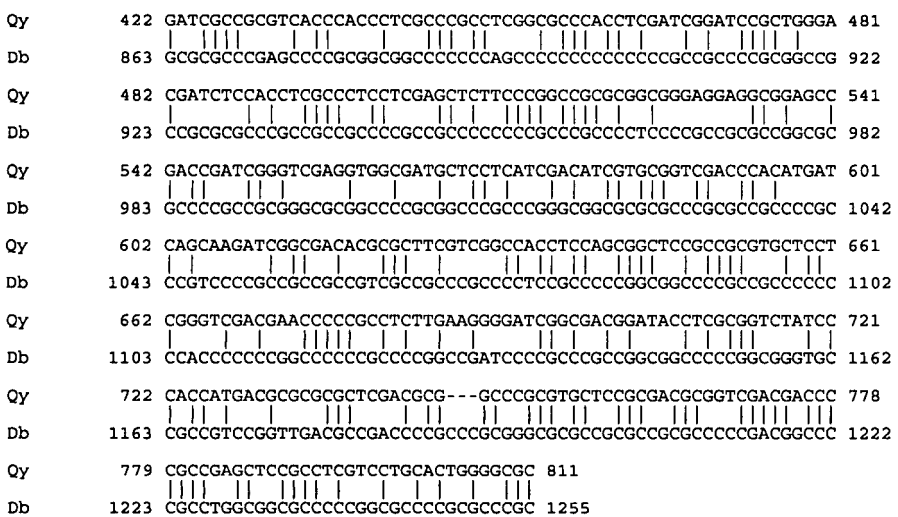
Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
Db        563 CCCCCCCGCCGCCGCCGGGCCCGCGCGCCCGCCGCGCCCCCGCCCCGCCGGCCCCCGCGG 622

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
Db        623 CGCCCCCCCGCCGGTCCCTTGCCTTGCCCCCCGGCGCCGTGCGCACCCCCCCGCGTCGCG 682

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
Db        683 CCTGGGGGGCGGCGCCCCCCCCGAGTGGCCCGGTCTTTTCTTTGCGCTCCCCGCCGCCCG 742

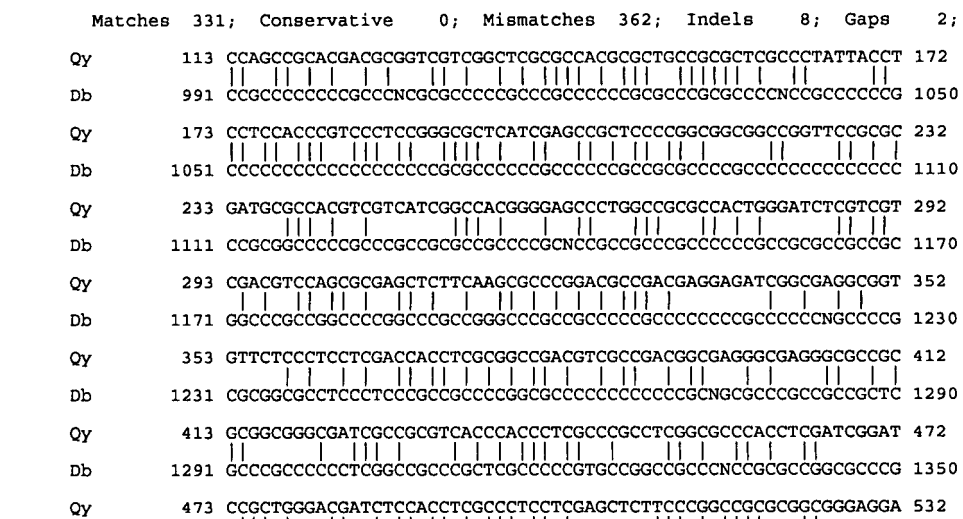
Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
Db        743 TGACCCGCCCCCCCCCCCCCCCCCCCCCCGCAGGGCGGGGGGGGGGGCCGCCTTCCCGCC 802

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
Db        803 CCCCCGCGCCCGCGCGCCCGCCCCGCGCGCCCCCGCACCCCCCCGCGTCGCGCGCCCGCG 862
```



```
RESULT 7
CC219595
LOCUS       CC219595               1785 bp    DNA     linear   GSS 12-MAY-2003
DEFINITION  CH261-62D17_RM1.2 CH261 Gallus gallus genomic clone CH261-62D17,
            genomic survey sequence.
ACCESSION   CC219595
VERSION     CC219595.1  GI:30541991
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1785)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 521
            High quality sequence stop: 576.
FEATURES             Location/Qualifiers
     source          1. .1785
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-62D17"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match            10.1%;  Score 90.6;  DB 9;  Length 1785;
  Best Local Similarity  47.2%;  Pred. No. 7.9e-07;
```



```
  Matches  331;  Conservative    0;  Mismatches  362;  Indels    8;  Gaps    2;
```



```
RESULT 8
AG171124
LOCUS       AG171124               1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                Sequencing: TJ
             LIBRARY
                Vector    : pBACe3.6
                R.Site 1  : EcoRI
                R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match             10.0%;  Score 90.2;  DB 10;  Length 1798;
  Best Local Similarity   42.8%;  Pred. No. 9.3e-07;
  Matches  378;  Conservative    0;  Mismatches  496;  Indels    9;  Gaps    4;

Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
              ||  |   |  ||||  || ||     |  |||   |||  || |  || | || |||
Db        754 CGNCGCGCNCGCGCGCGCCGGNCGCCCGCCGCGGCGCCCGCNCNCCGCGCGCCGCCGCGG 813

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
              ||    |  || ||      || ||  ||         | |  ||||   | |  | |||
Db        814 CGCNNNGCGCGCCGCGCNCGCGCCGCGCGNCGCCCCCGCGGCGCGCNCGCCCGCGNCGCC 873

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
                ||||  || ||   |||||  | |||  ||||||   ||       || |  || ||
Db        874 GCGCGCGCNCGCCGCGCCGCGCGCCNCGCGCCCGCGCGNNCCGCGGCGCCGCGCCCGCCN 933

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
               ||| |  ||||| |  ||   | |  | |  |      | |   ||| ||  || |
Db        934 CGCCCGCNCGGCGCNCNNCGCCGCCCCGCGCCNCNCGCCGCCGCGGCGCNCGCCGCCCGC 993

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
                |  |  | |||   ||    || |       | |  |   |       || || |
Db        994 GCGCCGCCGCCGCCCNCGGCGCGCCNGCNGCNCGCCCGCGCGCCNNNNNNGCGCCGGCGC 1053

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
              |||||  ||     |  |||| ||     |   |  |       |  || | ||      |  |
Db       1054 GCGCGCGC-GNGCGCGCGCGCGCNNNCGCGCCCGCGNCNCNCCGCGNCGCCCGCCCGCGC 1112

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
               | || || |    || |   || |  |  ||      |    | |           ||| ||
Db       1113 GCGCGNCCGCGGGCCGCCNCNCGCCCGCCGCGCNNCCGCNCNGCGGCGGCGCGCGCGC 1172

Qy        422 GATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACC----TCGATCGGATCCGCT 477
              |  || ||   | | |    | ||| |||  |  ||||| ||     ||      |   |||
Db       1173 GCGCGNCGGCGCGCGCCGGCGCGCNCGCGGCNNCGCCCGCCGGCGCCGCCGCGCCGCGCG 1232

Qy        478 GGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGG 537
              | | ||  | |   | ||| | |  |  ||    | ||  | |||    |  ||  |||
Db       1233 GCGNCGCGCGCGGGCGCGCNCGCGCCCGGCNGNGCGCGCNCCCGCNCGCGCCGGCCGCGC 1292

Qy        538 AGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACA 597
               |||  ||  || | || |   |||       |    ||     ||   |  ||   | |
Db       1293 CCCCGCGCGCNCGCGGCGCGCGCGCGCGCGGGCCGCGCGCGCGCG---GCGCGGGGCGCC 1349

Qy        598 TGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGC 657
              |  |     |  ||  |   |||||  || ||        | ||  |  |||||||  ||
Db       1350 GGCGCGCGCGGCGCGCGGCGCCGCGCGGCGCCGCGGGGNCGCCGCNCCGGCGCCGCGCGC 1409

Qy        658 TCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCT 717
               || | |  |   |  | | |       |     |  |      || |     |  |||  |
Db       1410 GCCNCCGCGCNCGGNCCGCGGCGGNCGNCCGGCCGCCNCNCGGCNCGCGCGGCGCCGCGG 1469

Qy        718 ATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACC 777
              | | |   | |||      | |  | |  | |  ||| ||  | |  | ||  |  ||  |
Db       1470 CGCGCGCGCCGGCGCCNCGCCGCNCCNCNCCGCGCCCGCNNCNCNNC-CGNNCCGCGCGC 1528

Qy        778 CCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGCGGATGAGCGATCACGACGTGA 837
              |  |    |  || | ||     | | | |  |||| |||   | ||   |    |  |
Db       1529 CNNCGCGNCCGCGNCNCGCNNCNCNCCGCGCGGCGGCGCCGGCGCCGGCGCCGCCCGGN 1588

Qy        838 TCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCGGTCAC 880
               ||   ||     |  || || |  |  |  ||  |    | |
Db       1589 CCGCNCGGCGCNCCGCCCCGCGCGNCNCGCGCGNCGNNCGCGC 1631


RESULT 9
CL078342
LOCUS       CL078342                1674 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-149D22_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-149D22, genomic survey sequence.
ACCESSION   CL078342
VERSION     CL078342.1  GI:40534255
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1674)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 300
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1. .1674
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-149D22"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             10.0%;  Score 90;  DB 10;  Length 1674;
  Best Local Similarity   46.6%;  Pred. No. 1e-06;
  Matches  346;  Conservative    0;  Mismatches  394;  Indels    2;  Gaps    2;

Qy         71 CGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGGT 130
              | ||| ||  |  ||| ||   ||       ||  | |  || || ||| | |  |  |
Db       714 CCGCGCCGCGCGCGCGCCCCGCCGCCCCCCGCCCCCCCCCCCCCCGCCCCGCCCCCGGCC 773

Qy        131 CGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCCG 190
              |  || | ||| || |||||  || | | |||||       | || ||  ||  ||| ||
Db       774 CCGCGCCCCGC-CCGCGCGCCCCCCCCCCCGCCCCCGCCGCCCCCCCCCGCCCCCCCGCCC 832

Qy        191 GGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCAT 250
               |   |   |  || | ||||| || |  |||   |  |||    | | |  |  |
Db       833 CCCCGCCCCGGCCCCCCCCCCCGCCGCCCCCCGCGCCGCCGCCCGCCCCGCCCGCCCCG 892
```

```
Qy        251 CGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGCT 310
Db        893 CGCCCCCCCCGCCCCCCCCCCCCCGCCCGCCCCCCCCCCCCCCCGCCCCCCGCCCCCCC 952

Qy        311 CTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACCA 370
Db        953 CCGCGCCCCCCGCGCCCGCCCCGCGCGCGCCGCCCCCGCCCCGCCCGCCCCCCGCGCCCC 1012

Qy        371 CCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGC 430
Db       1013 CCCCCCCGCCCCCCGCGCCCCCGCCCCCGCCCCCCCCGCCCGCGCCCCCGCCCCCCCCG 1072

Qy        431 GTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCCA 490
Db       1073 CCCCCCGCCCCCCCCCCCCCGCCCCGCGCCCCCCCCCCCGCCCCCGCGGCCCGGCCCGCC 1132

Qy        491 CCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCG 550
Db       1133 GCCCCCCCCGCCCGCCGCCCCGCCCGCCCCCGCGCCCCCGCCCGCCGCCGCGCCCGCCCC 1192

Qy        551 GGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGAT 610
Db       1193 NCCCGCGCCCCCCCGCGCCCGCCGCGCCCGGCGCCCCCCCCCGCCGCGCGCCCGGCCCCCC 1252

Qy        611 CGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGAC 670
Db       1253 CGCCCCGCCGCCCGCGCGCCCCCCCCCGCGCCCGCCCCGCCCCGCCGCCCCCGCCCCGCC 1312

Qy        671 GAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGAC 730
Db       1313 CCCCCGCCGCCGCCCCCCGGCCCCCCGCCCCGCCCGCC-CGCGCCCCCCGGCCCCCNCCC 1371

Qy        731 GCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCG 790
Db       1372 CCCCGCCCGCGNCGCCGCCCCGCGCCGCCCGCCGCCCGGCCCCGCCCCCGCCGCCGCCC 1431

Qy        791 CCTCGTCCTGCACTGGGGCGCG 812
Db       1432 CCGCCCCCCNCGGCGCCGCCCG 1453


RESULT 10
AG058789
LOCUS       AG058789               1006 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-045M14.F, genomic survey sequence.
ACCESSION   AG058789
VERSION     AG058789.1  GI:16596250
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1006)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
            LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1006
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-045M14.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             10.0%;  Score 89.8;  DB 10;  Length 1006;
  Best Local Similarity   44.3%;  Pred. No. 1.2e-06;
  Matches  357;  Conservative    0;  Mismatches  445;  Indels    3;  Gaps    2;

Qy         98 GGTGCAGCGCTTCCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCG 157
Db        138 GGNGCGCCGGCGGCGCGGGNGCCCCCGCGGGNGCGCGGGCGGGCGGGNGGCGCGGCGGGC 197

Qy        158 CTCGCCCTATTACCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGG 217
Db        198 CGGCGCGCCCCGCGGCCCGCGGCGCGCGCGCGGGGGGCGAGAGAGGCCTCGCAACGAGGC 257

Qy        218 CGGCCGGTTCCGCGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCA 277
Db        258 GGGCCGCGCCGCCGGCNCCGCGGCGCCCGCACAACCACAAAACCACGCCCNAAAANGCCC 317

Qy        278 CTGGGATCTCGTCGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGA 337
Db        318 GCGGCCCTGCNCCGNGTTGTTCNNNGCCGCCCGCGGGCGCCCCCGCCGCGCCGCCGGGN 377

Qy        338 GATCGGCGAGGCGGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGA 397
Db        378 GNGCGGCGCGNCCGCCCGCGCCGCGGGCGGCCGCCCCGNCGCCGNCCNCNNGGGGCGCGG 437

Qy        398 GGGCGAGGGCGCCGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGC 457
Db        438 GGCCCGCCCNGGCGCGGNGCGNGCGGCCCGCGCCCCCCGCGCCCGCCCCCGCCCCCGCCC 497

Qy        458 CCAC--CTCGATCGGATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCG 515
Db        498 CCCCGGCCCCGACGCGGCCGCCCCCGCGCGCGCGCCGCCGCGCGCGGCGGGCGGGCGGCG 557

Qy        516 GCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATC 575
Db        558 GCGCCCGGGCGNCNNNGCGCCCCCGNCNCCGCGCGGCGCGCCGCCGCCGCGCGCCGGCGC 617

Qy        576 GACATCGTGCGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCAC 635
Db        618 CCCGCCGCCCGCGCGCCCGCGCGCCGCCGCCGGCCCGCGCCCGCGCGCCGCG-CGGCCGC 676

Qy        636 CTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATC 695
Db        677 CCCGCGCGGCCCCGGNCCCCGCCCCCCGGNNGCNGCCGCGCCGCGCCGCCGCGGCGG 736

Qy        696 GGCGACGGATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTG 755
Db        737 CCCGCGCGGCGGCCCGGCCGCCGCCGCCCCCGCGGNCGCGCCCCCGGCCCGCCCCCGCCC 796

Qy        756 CTCCGCGACGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTG 815
Db        797 CCCCGCCCCGCNGCCGNCCCGCCCGCGCGGACCCGCGCCGGGCGACCCNCCGCCGCCCCC 856

Qy        816 CGGATGAGCGATCACGACGTGATCGGCAGGGAGGTCCACCGGCTCTTCCGGATCGAGGCG 875
Db        857 GCGCCGCCGCCGCCGGCCCGCCGCCGCGGGGCGGCCCGCCGGCGGACCCCGGCCGACGGA 916

Qy        876 GTCACCGAGGAGGATCGCGCCGCGG 900
```

```
Db        917 GCCCCCCCCCCCCCGCGCGNCCCCG 941


RESULT 11
CL498921/c
LOCUS       CL498921               1319 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_661_F04.v2 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_661_F04.v2, genomic survey sequence.
ACCESSION   CL498921
VERSION     CL498921.1  GI:45995121
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1319)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS828795; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1319
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_661_F04.v2"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             10.0%;  Score 89.8;  DB 10;  Length 1319;
  Best Local Similarity   46.5%;  Pred. No. 1.1e-06;
  Matches  343;  Conservative    0;  Mismatches  390;  Indels    4;  Gaps    2;

Qy          59 GCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCC 118
               || ||  | |  | ||||     |||    || ||      |       ||  ||||| |||
Db         905 GCGCGCCGCGCGCNGCGGGCCGGCGGGTCTCGCGCCCCCCGGCGCCGGCCCCCGCCCGCC 846

Qy         119 GCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCA 178
               || || | |||    |  | ||   | ||    || |||| ||||       | ||  |
Db         845 GCCCGCCCCGGCGCCCCCCGCGGCGCGCGGCGCGCGGCGCCCGCCGCGGCGGCGCCCGCC 786

Qy         179 CCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCG 238
               |||  | | || | |||    ||  |||  | ||  ||||   ||     | |||  |||
Db         785 CCCCGCGCGCCCGCCGCCGCGCGGCCCGGCCGCCCCCGGCCCGCGCGCGCCGGCGCGGCG 726

Qy         239 CCACGTCGTCATCGGCCACGGGGAGCCCT---GGCCGCGCCACTGGGATCT-CGTCGTCG 294
               || || || |  |   |   ||   |||    |||||||| | | |  |  ||  | ||
Db         725 CCCCGCCGCCGCCCCGCCGCGGCGCCCCGGCGGGCCGCGCCCCGGCGCGCGGCGCGGCCC 666

Qy         295 ACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGT 354
                 || || ||||   | |  |  |||| ||    |   ||  |    ||      ||
Db         665 GCGCCCCGCGCCCCCCCCCCCCGCGCGCGCGGCCGCGCGCCGGTCCCGCGCCCGCCCCCC 606

Qy         355 TCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGC 414
                 | ||| |  ||    ||  |    |  |  | ||  || |  |  |      |||||||

Db         605 CCGCCCGCGCCGCGGCCCCGGGCCGCCGCCGCCCCCGCGCCCCGCCCTGCCCCGCCGCGG 546

Qy         415 GGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCC 474
               |||   |  |||  || | |||  |   || |||  ||    |     ||   |   |
Db         545 CGCGCCGGCCCGCGCCGCCCCCCGCTCGGCGCGCGCCGCGCGCGCGGGCGNNGGCGCCG 486

Qy         475 GCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGG 534
               |  |||         | || ||| |  | |  ||    | ||||||||| ||   ||  |
Db         485 GGGGGGGGCCGGCGCGCCGCGCGCGGCCCCCGCGGCGCGCGGCCGCGCGCCGCCCGGGCG 426

Qy         535 CGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCC 594
               || ||||  |   |||| |       |    | ||    |  |  ||  ||  || |||
Db         425 GGGCGCCGGGCCCCCGGGGCCCCCGCCCCCGCCCCCCGCCCCCCCGCGCCCGCCCGCCCC 366

Qy         595 ACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCG 654
                  |  |  ||||||  | | |  |  |||  |   ||  | |||||||
Db         365 CGCCCCGCCCCCGTGTCGGCGGGCCCCCCCCNGGGGGCGCGCCGGCGCCCCCCCGCCGCC 306

Qy         655 TGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGG 714
               |    |||  | |  |||||||  |       |   || |   |    |   ||||
Db         305 GGGGGGGCGGCGCCCCCCCCCCCCGCGGCCCCCGTNGNNGCGCCCGGGCCGGCGGGGCGG 246

Qy         715 TCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACG 774
               |  ||  ||    |  | | | ||  |  ||| |  || | |||  | ||  | |
Db         245 CCGCCCCGCCCCGCCCCGCCCCCCGCGTTTCCCCCCCCCCGCCCGGCGGTGTGGNGGCCC 186

Qy         775 ACCCCGCCGAGCTCCGC 791
                 ||  | |   | ||||
Db         185 CCCGGGGCCGCCCCCGC 169


RESULT 12
CL500927
LOCUS       CL500927               1474 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_68_H09.v3 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_68_H09.v3, genomic survey sequence.
ACCESSION   CL500927
VERSION     CL500927.1  GI:45998247
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1474)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS803245; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1474
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_68_H09.v3"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
```

ORIGIN

Query Match 9.9%; Score 89.6; DB 10; Length 1474;
Best Local Similarity 48.4%; Pred. No. 1.2e-06;
Matches 355; Conservative 0; Mismatches 375; Indels 4; Gaps 4;

```
Qy         50 GTGCGGCCTGCTCGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTT 109
Db         97 GGGGGGGGGGGTTTTGGGGGGCGGGGCGGCGGCGGGGGCCCCCGGCCGCGCGCCGCGCCC 156

Qy        110 CCGCCAGCCGCACGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTA 169
Db        157 GCGCCGGCCGCCCCGC-CGGGCGGCCGGCGGCGGCGCCCCCGCCCGCCCCCCCCCGCCCC 215

Qy        170 CCTCCTCCACCCGTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCG 229
Db        216 CCCCCCGCGCGCGCGCGCGCGCGCGCCCCCCCCGGCGCCGCGCCGCGGGGGGCGCGCGCG 275

Qy        230 CGCGATGCGCCACGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGT 289
Db        276 CGGGCCGNATTCTTTTTTCTTTTCGCCGGGCGGCGCGGCGCGGGGGCGCGGGCGGCGCGC 335

Qy        290 CGTCGACGTCCAGCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGC 349
Db        336 GGCGCGCGGCC-GCCCGGGCGGCGCCCGGGCCGCCGCGGGCCCCGGGGGGCGGGCGGGNC 394

Qy        350 GGTGTTCTCCCTCCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGC 409
Db        395 GGCGCGCGCGGGGCGGGGCGGCGCGGC-GCCGGCGCCGCCGGCCGGCGGGCCGCGGCCGC 453

Qy        410 CGCGCGGCGGGCGATCGCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCG 469
Db        454 CGCGCGGCGCGCGCGCGCGGCGCGNCCCGCGCGGCCGCCGCGCGCCCCGCGCGCCCGGCG 513

Qy        470 GATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGA 529
Db        514 GCCCCGCCGCCGCGCGCCGCCCCGCCGCCGCGCCG-GCGCGCGCGCGCGCCCGGCGGNCCGC 572

Qy        530 GGAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTC 589
Db        573 GCGGCCGGCGCCGGGCGCGCCGCGCGGCGGCGCCGGGCGCGGNGCGCCCCCGCGCGGCGC 632

Qy        590 GACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCG 649
Db        633 GGCCCCCGGCGGGCCGCGGGCGGGNGCGGGCCGCCCGGCCCGCGCGCGCCGTCGCCGCGC 692

Qy        650 CCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCT 709
Db        693 GCGCCGACGGCGGCGCGGCGCGCGCCGCGGCGGGGCGGCGGCCCGCCGCGCCGGCGCGGG 752

Qy        710 CGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGT 769
Db        753 CGGTGGCCGGCGGGGCGGGTCGCGCCCGCGCGGCGCGCGGCGCGGACGCGGCGGTCCGCG 812

Qy        770 CGACGACCCCGCCG 783
Db        813 CCCGCGCCCGCGGCG 826
```

RESULT 13
BG810038/c
LOCUS BG810038 1116 bp mRNA linear EST 22-MAY-2001
DEFINITION mgct002xi01f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe grisea cDNA clone mgct002xi01f 5', mRNA sequence.
ACCESSION BG810038
VERSION BG810038.1 GI:14181018
KEYWORDS EST.
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Choi,W. and Dean,R.A.
TITLE Construction and sequence analysis of an appressorium stage cDNA library in the rice blast fungus, Magnaporthe grisea
JOURNAL Unpublished (2001)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu
Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES Location/Qualifiers
source 1. .1116
/organism="Magnaporthe grisea"
/mol_type="mRNA"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgct002xi01f"
/dev_stage="Germinated conidia on apprressorium-inductive surface"
/clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
/note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI; Site_2: XhoI; The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library contains over 55,000 clones with average insert size of 1.5 kbp."

ORIGIN

Query Match 9.9%; Score 89.4; DB 2; Length 1116;
Best Local Similarity 43.1%; Pred. No. 1.4e-06;
Matches 359; Conservative 0; Mismatches 456; Indels 18; Gaps 2;

```
Qy          2 CGGCCGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCT 61
Db        973 CGGGCGGGCCGCGGGCCGGTGCCGCGCCGCCGGCCGCCCCTCTTTCCGGCCGCGGCGGCC 914

Qy         62 CGGGGTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCA 121
Db        913 CGCGGGGCCCCTGCGCGCGGGGGGCGCGCGCGGCNCCGGGCGCCCCNGCGCCCCCCCCGC 854

Qy        122 CGACGCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCC 181
Db        853 GNCCCCGCGCCGCGCNGCCNCCCCCCCCCGCGCCCCCCGGNGGCGCGCCGCCCGCGCCCC 794

Qy        182 GTCCCTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCA 241
Db        793 GCCCCCGCGNCGCNCCCGCCGNGCGCGCCGCGCCGGCCCCCGCCGCCCGGCCGGCCGGCG 734

Qy        242 CGTCGTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCA 301
Db        733 CGCCCCCGCCGCCCCCGCGGCGCCGCCCCCCGCCGCGCGGCCGCGGGGCGCGCGGGGG 674

Qy        302 GCGCGAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCT 361
Db        673 GCGCGCGGGGGCCGCGCCCGCGGGCGCGCGGCGGCCGCGCGCCGGCNCGCGCGCGCCGCC 614

Qy        362 CCTCGACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGC 421
Db        613 CCCCGCCGGGGCGCCGCGCCGCCGCGCCGGGCGCGGCGGCGCGCGGGCCCGCCTCCGCGC 554

Qy        422 GATCGCCGCGTCACCCAC----------CCTCGCCCGCCTCGGCGCCCACCTCGATCGG 470
Db        553 GGCCGCCGGCGCGGCCGCGCCGCGCCGCCCCGGCCGCGCGGCGGCGCCCCCGCCGG 494

Qy        471 ATCCGCTGGGACGATCTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAG 530
Db        493 CCGCCCCGGCGCCCGGCCCCCNNCGGCCCGGCNCCCGCNCCCCCGCCGCGGCGCCGCGCG 434

Qy        531 GAGGCGGAGCCGACCGATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTG------ 584
Db        433 GCGGCGGCGCCNNCCCCGCCCCCCCCCNNCCCCNNCCCCCGCCCCCCCCCCNNGCCCCGC 374

Qy        585 -CGGTCGACCCACATGATCAGCAAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCG 643
```

```
Db        373 CCGNNCGGCGCCCGGCCCCGCCCGCGCCGGCGNGCCGCCGGGGGNGGCCCCCCCCCCGCC 314

Qy        644 GCTCCGCCGCGTGCTCCTCGGGTCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGG 703

Db        313 GCCCCGCCGGNCGCGGCCCGGCCCCCCCGGCGCCCGCGGCGCGGGCGCCNNNNGGGGNGG 254

Qy        704 ATACCTCGCGGTCTATCCCACCATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGA 763

Db        253 CCCCCCCCCCCCCCNNCCCCCNNNNNNCCCCCCNCNCNNNGCGGNCCNCCCCCNNCNCNN 194

Qy        764 CGCGGTCGACGACCCCGCCGAGCTCCGCCTCGTCCTGCACTGGGGCGCGGTGC 816

Db        193 NANNNNNNNNNCNCCCCCCCCCCCCCCCCNCCCNCCCNTTTTGTNCNNNGGTTC 141


RESULT 14
BZ569478/c
LOCUS       BZ569478               1116 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_8391.x2r5p12 pacs2-164 Pseudomonas aeruginosa genomic
            clone pacs2-164_8391, genomic survey sequence.
ACCESSION   BZ569478
VERSION     BZ569478.1  GI:27204164
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1116)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1116
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8391"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             9.9%;  Score 89.4;  DB 9;  Length 1116;
  Best Local Similarity   49.2%;  Pred. No. 1.4e-06;
  Matches  366;  Conservative    0;  Mismatches  368;  Indels   10;  Gaps    5;

Qy         70 TCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGACGCGG 129

Db       1095 TCGGCGCCGCCCCGGCCCCCGGCCGCCGGCCGCGGCCCGGCCCGGCGCCCCGCCGCGCGC 1036

Qy        130 TCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCCCTCC 189

Db       1035 GGGCGGCCCCGCGGGGCGCGC-GCCACGCGNCCGCCGGCCCGCCGGCCACGACGCCGCGC 977

Qy        190 GGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTCGTCA 249

Db        976 GGGGNCGCGGCCCGCGGCCCGCCGG-GCCGGCGGGGGGCGGGGGCGCCCGCGCGGCGCGC 918

Qy        250 TCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGCGAGC 309

Db        917 GCGCGGGCGCCGCGGCCGCCCGCGGCCCCGCGCCCCCCGCGCGGGGCGGCGGGCGCCGCC 858
```

```
Qy        310 TCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTCGACC 369

Db        857 CCGCCCCGCGGCCGCGCCG------GGGGCGCGGCCCGGCGCGGCCGCCCGACCCGCCCC 804

Qy        370 ACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATCGCCG 429

Db        803 CCCCCGCGGGGGCCCGGGGGGCGGGCGGGGGCGGGCGCCCCCCGCCCCCGGGGCGCGCCC 744

Qy        430 CGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGATCTCC 489

Db        743 GCGCGCGCCGCGCCGGGCCCCGCGGCCCCCGCCGCCGGGCCGGGCCCCCGCGGCGCCGCG 684

Qy        490 ACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACCGATC 549

Db        683 GGGCGGCGCCGCGCGCCCCGGGCGCCGGCGCCNGCCCGCCCCCCGCGG-GCGCGCCCCGC 625

Qy        550 GGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGCAAGA 609

Db        624 CCGCCCACGCGGCCCGGCGCCCGCGCCCCGGGCGGCGCCGCGCGGGGCGGCCNCCGCGC 565

Qy        610 TCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGGTCGA 669

Db        564 GGCCCGCGCNGCGCCCCGGCGGCGGCCCCCGGCGGGCCGGGCGGCGGCCGCCCGGGCCCC 505

Qy        670 CGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACCATGA 729

Db        504 CGCCGCGGCGCGCCCCGGGGGGCGGCGGC-CCGCGGCGCGGGCGGCGCCCCGCCCCGCGC 446

Qy        730 CGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAGCTCC 789

Db        445 CGCGCGGCCGCCNCGGCGGCCGCGCGCGGGCCCGCGCGCCCCNCGCGCCCGCCCCGCCGC 386

Qy        790 GCCTCGTCCTGCACTGGGGCGCGG 813

Db        385 GCGCCGCCGCCCCCGCGGCCGCGG 362


RESULT 15
CL090560/c
LOCUS       CL090560               1821 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-17N15_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-17N15,
            genomic survey sequence.
ACCESSION   CL090560
VERSION     CL090560.1  GI:40584195
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1821)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 591
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1. .1821
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-17N15"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
```

Library Segment 1"

ORIGIN

Query Match 9.9%; Score 89.4; DB 10; Length 1821;
Best Local Similarity 46.4%; Pred. No. 1.3e-06;
Matches 365; Conservative 0; Mismatches 416; Indels 6; Gaps 2;

```
Qy          6 CGGGACGCGGACGACGTCTATCACCCGTTCCGCGAGCTCTTCGAGTGCGGCCTGCTCGGG 65
              || |  |||| ||        | || |   ||||    |       || |   || || |
Db       1627 CGCGCGGCGGCCGCTCCGCGCCCCCGGGCGCGCGCCGCCCGGCGCCGCCGGGCGCGCGCG 1568

Qy         66 GTGATCGGCGGCGATCCGGCGGCCGAGCGGAAGGTGCAGCGCTTCCGCCAGCCGCACGAC 125
                | ||  |   ||| ||  || |     |  ||  | |  |  |   |  |  | |
Db       1567 CGCCCCCGCCACCGCCCGCCGCGCGCGGCCCCGCGGCGCCCCGCCGCCGGCCGCCCGGGC 1508

Qy        126 GCGGTCGTCGGCTCGCGCCACGCGCTGCCGCGCTCGCCCTATTACCTCCTCCACCCGTCC 185
              || | || | || ||  |    ||| ||||| | ||| |       |  | || ||   |
Db       1507 GCCGGCGCCCGCGCGGCCGGGCCGCGGCCGCCCGCGCGCGCGGCGCCGCGCCGCCGCGCG 1448

Qy        186 CTCCGGGCGCTCATCGAGCCGCTCCCCGGCGGCGGCCGGTTCCGCGCGATGCGCCACGTC 245
              | ||   | | |   | | ||| | ||||  |  | |      ||| |  |||   || |
Db       1447 CACCCCCCCCGCCGGGCGGCGCGCGCCGGGCGGCGGCCCGGGGGCGGGGGGCGGGCCGCC 1388

Qy        246 GTCATCGGCCACGGGGAGCCCTGGCCGCGCCACTGGGATCTCGTCGTCGACGTCCAGCGC 305
              |    |  |  | |   |||| |||    || | |       |  || |       | || |
Db       1387 GGGCGCCACGCCCGCCCGCCCGGGCGCGCCCGCCGCCGCGCCCGCGCCCCGCCGCGGCCC 1328

Qy        306 GAGCTCTTCAAGCGCCCGGACGCCGACGAGGAGATCGGCGAGGCGGTGTTCTCCCTCCTC 365
              |       |  ||| |||      |  || || |  ||    || |      |     || |
Db       1327 GCCACGGGCGCGCGGCCGCCGCGCCGCGGGGGGCCCGCGCGGGGGCGCCGCAGGGCCCCC 1268

Qy        366 GACCACCTCGCGGCCGACGTCGCCGACGGCGAGGGCGAGGGCGCCGCGCGGCGGGCGATC 425
              |  |  | ||||||   || |  |  | ||| ||||  || |||    || || ||    |
Db       1267 GGGCGGCCCGCGGCG--CGGCCGCCGCCGCGCGGGCCCGGCCGCGCGGCCGCCGGGCGCC 1210

Qy        426 GCCGCGTCACCCACCCTCGCCCGCCTCGGCGCCCACCTCGATCGGATCCGCTGGGACGAT 485
              |||||  | |||   | | | ||||   | |||  || ||  | |   ||    | ||
Db       1209 GCCGCCCCGCCCCGGCCCCCGCGCCCGCGGGCCGGCCGCGCGCCGGCGCGGCCCGCCGGC 1150

Qy        486 CTCCACCTCGCCCTCCTCGAGCTCTTCCCGGCCGCGCGGCGGGAGGAGGCGGAGCCGACC 545
              | || |  ||  | |  || || |  |  |||||  | ||  |  | | |   ||   ||
Db       1149 CGCCCCGGCGGGCCCGCCGGGCGCCGCGGGGCCGGCCCGCCCGGCGGGCCCCCGCGCGCC 1090

Qy        546 GATCGGGTCGAGGTGGCGATGCTCCTCATCGACATCGTGCGGTCGACCCACATGATCAGC 605
              |    || ||  | | |    | |  |  || |  |  |  |     || |     | ||
Db       1089 GCCGCGGCCGCCGGGCCCCGCCGCTCCGCCGCCCGCCGGGCGCGCGGCCGCCGCGGCGGC 1030

Qy        606 AAGATCGGCGACACGCGCTTCGTCGGCCACCTCCAGCGGCTCCGCCGCGTGCTCCTCGGG 665
                |||||    |||||  |  ||||| || ||    |   |||||||    || ||
Db       1029 GCCGGCGGCCCGCCGCGCCGCCGCGGCCCCCCCCCCCGCGGGGCGCCGCGCCGCCCCCGCC 970

Qy        666 TCGACGAACCCCCGCCTCTTGAAGGGGATCGGCGACGGATACCTCGCGGTCTATCCCACC 725
                |    ||||| ||  |  |          ||  |    || ||||       |   ||
Db        969 GGCCGG----CCCGCGTCGCGCGGCCCCCGCCCGCGGCCCCCCCCGCGCCGCCGCGGGCC 914

Qy        726 ATGACGCGCGCGCTCGACGCGGCCCGCGTGCTCCGCGACGCGGTCGACGACCCCGCCGAG 785
                ||||| ||| || || ||||||||  | | ||   |  | ||  | ||    |  |
Db        913 GGCGCGCGCCCGCGCGCCGGGGCCCGCGGTCGCGGCCGGGGCGCCGGGGCCCGGTGCCCG 854

Qy        786 CTCCGCC 792
              | |||||
Db        853 CGCCGCC 847
```

Search completed: April 28, 2006, 12:17:51
Job time : 4467.48 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 434.473 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_76000_76900
Perfect score: 901
Sequence: 1 ggcgggggccttccggctgg.....cgggcgcggcccacgcgtcg 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		901	100.0	76994	12	ADP64454	Adp64454 Sorangium
2		84.4	9.4	114955	2	AAX53491	Aax53491 Human ade
c 3		81.2	9.0	11604	10	ADD26442	Add26442 Streptomy
c 4		81.2	9.0	11604	10	ADG47787	Adg47787 Streptomy
5		80.2	8.9	3024	14	AEB57391	Aeb57391 Silk elas
c 6		80	8.9	1722	4	AAD14503	Aad14503 S. clavul
c 7		80	8.9	1722	10	ADD26444	Add26444 Streptomy
c 8		80	8.9	1722	10	ADG47789	Adg47789 Streptomy
c 9		80	8.9	1722	10	ACA62926	Aca62926 DNA encod
c 10		80	8.9	11604	4	AAD14501	Aad14501 Streptomy
c 11		80	8.9	11604	10	ACA62924	Aca62924 Clavulani
c 12		80	8.9	15079	4	AAD14499	Aad14499 Streptomy
c 13		80	8.9	15079	10	ACA62922	Aca62922 Streptomy
c 14		80	8.9	15120	10	ADD26455	Add26455 Streptomy
c 15		80	8.9	15120	10	ADG47775	Adg47775 Streptomy
c 16		80	8.9	29870	8	AAD36874	Aad36874 Streptomy
17		78	8.7	9222	8	AAL61171	Aal61171 Actinosyn
c 18		78	8.7	15079	2	AAQ91580	Aaq91580 S. clavul
c 19		78	8.7	82746	8	AAL61224	Aal61224 Actinosyn
c 20		77.8	8.6	5223	8	ACA26233	Aca26233 Prokaryot
c 21		77.8	8.6	9222	8	AAL61171	Aal61171 Actinosyn
22		77.8	8.6	82746	8	AAL61224	Aal61224 Actinosyn
c 23		77.4	8.6	1431	8	ABZ37535	Abz37535 Streptomy
24		77.4	8.6	59816	8	ABZ37516	Abz37516 Streptomy
c 25		77.4	8.6	59816	8	ABZ37515	Abz37515 Streptomy
c 26		77.2	8.6	58857	3	AAA58471	Aaa58471 Nucleotid
c 27		75.8	8.4	1680	1	AAN81334	Aan81334 Complete
c 28		75.8	8.4	1722	1	AAN50035	Aan50035 Sequence
c 29		75.4	8.4	4770	8	AAL61172	Aal61172 Actinosyn
c 30		75.2	8.3	11220	3	AAZ87298	Aaz87298 S. venezu
c 31		75.2	8.3	11220	12	ADL91917	Adl91917 Streptomy
c 32		75.2	8.3	36778	3	AAZ87318	Aaz87318 S. venezu
c 33		75.2	8.3	36778	12	ADL91933	Adl91933 Streptomy
c 34		75.2	8.3	37948	3	AAZ87285	Aaz87285 S. venezu
c 35		75.2	8.3	38506	3	AAA75633	Aaa75633 Nucleotid
c 36		75.2	8.3	38506	3	AAZ56001	Aaz56001 Recombina
c 37		75.2	8.3	38506	8	ADA09418	Ada09418 Cosmid pK
c 38		75.2	8.3	38506	10	ADH53462	Adh53462 S. venezu
c 39		74.8	8.3	2118	8	ACA26211	Aca26211 Prokaryot
c 40		74.8	8.3	22533	13	ADQ91713	Adq91713 Polyketid
c 41		74.8	8.3	22533	13	AEB87005	Aeb87005 Streptomy
c 42		74.8	8.3	22533	13	AEB86802	Aeb86802 Streptomy
c 43		74.8	8.3	164051	13	ADQ91695	Adq91695 Polyketid
c 44		74.8	8.3	164051	13	AEB86862	Aeb86862 Streptomy
c 45		74.8	8.3	164051	13	AEB86659	Aeb86659 Streptomy

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

```
PS   Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC   The present invention describes an isolated recombinant polynucleotide
CC   (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC   synthase (PKS) protein or a fragment comprising one or more domains of
CC   the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC   host cell (III) comprising (II), or (I) integrated into the cell
CC   chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC   disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC   the native disorazole PKS by the inactivation of one or more domain of
CC   the disorazole PKS and/or addition of one or more domain of a non-
CC   disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC   expression system (VI) capable of producing a disorazole synthase domain
CC   in a host cell, where the system comprises an encoding sequence for a
CC   disorazole polyketide synthase domain, and the encoding sequence is
CC   operably linked to control sequences effective in the cell to produce RNA
CC   that is translated into the domain; (7) host cell modified to contain
CC   (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC   in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC   eliminate production of disorazole; (9) an isolated polypeptide encoded
CC   by (I); (10) hybrid polyketide synthase comprising one or more
CC   polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC   disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC   of at least 200 base pairs with a sequence identical or substantially
CC   identical to a protein encoding region of the 77294 base pair sequence of
CC   SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC   inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC   (I) is useful for producing disorazole polyketides. (III) is useful for
CC   producing a polyketide, which involves growing (III) under conditions,
CC   where a polyketide synthesised by a PKS comprising a protein encoded by
CC   the polynucleotide molecule is produced in the cell. (I) is useful for
CC   producing libraries of PKSs, where the polyketide producing colonies are
CC   identified and isolated, and the produced polyketides are used
CC   collectively in a panel to represent a library or assessed individually
CC   for activity. The disorazole polyketide produced by (I) is useful in
CC   molecular biology, chemistry, recombinant DNA technology, human and
CC   veterinary medicine and agricultural applications. The disorazole
CC   polyketide produced by (I) is formulated as a pharmaceutical composition,
CC   useful for its activity against microorganisms, in a mammal. The present
CC   sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC   disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 901;  DB 12;  Length 76994;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-131;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75700 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 75759

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75760 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 75819

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75820 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 75879

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75880 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 75939

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75940 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 75999

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76000 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 76059

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76060 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 76119

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76120 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 76179

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76180 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 76239

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76240 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 76299

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76300 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 76359

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76360 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGC 76419

Qy        721 GGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76420 GGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCC 76479

Qy        781 GTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76480 GTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTC 76539

Qy        841 GCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76540 GCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTC 76599

Qy        901 G 901
              |
Db      76600 G 76600


RESULT 2
AAX53491
ID   AAX53491 standard; DNA; 114955 BP.
XX
AC   AAX53491;
XX
DT   05-JUL-1999  (first entry)
XX
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
XX
OS   Synthetic.
XX
PN   WO9913886-A1.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US019419.
XX
PR   17-SEP-1997;   97US-0059160P.
PR   09-JUN-1998;   98US-00093972.
XX
```

PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 1999-229400/19.
XX
PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT vasoconstriction.
XX
PS Disclosure; Page 37; 120pp; English.
XX
CC The specification describes antisense oligonucleotides (AAX52869-X55271)
CC directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene initiation
CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'
CC -end and the juxta-section between coding and non-coding regions and all
CC segments of RNAs encoding proteins associated with one or more diseases,
CC conditions or mixtures. The antisense oligonucleotides may be derived
CC from sequences AAX55272-74. These multiple target oligonucleotides
CC (specifically AAX55180-271) can be used for the antisense treatment of
CC diseases and conditions. Typical diseases and conditions are those
CC associated with impaired respiration and inflammation, including lung
CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,
CC acute asthma, allergies, asthma, impeded respiration, respiratory
CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,
CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary
CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
CC colon cancer, breast cancer, lung cancer, pancreatic cancer,
CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
CC well as all types of cancers which may metastasize or have metastasized
CC to the lungs, including breast and prostate cancer
XX
SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

Query Match 9.4%; Score 84.4; DB 2; Length 114955;
Best Local Similarity 32.0%; Pred. No. 9.2e-05;
Matches 263; Conservative 89; Mismatches 467; Indels 2; Gaps 1;

```
Qy        66 GCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCC 125
             ||||    || |:   :  |  |:|| |:  |||   || | ||  |||:   :   | |
Db    104292 GCCGGCCGGGCCSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCCGGGCSNNNDNNCCGC 104351

Qy       126 GGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGC 185
             :|    |||||  |||    | || |:   :  ||| :|  |: |            | |
Db    104352 BGGCCBGGGCGCGCCGCCGGCCGGGSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCCGG 104411

Qy       186 CGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGG 245
             :      |      | : || |||| |||  || :   :   || :|     |
Db    104412 SNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCCGSNNNDNNCCGCBGGCCBGGGCGCGCC 104471

Qy       246 CCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGC 305
             |  |||:   :     : |||: | ||  || | | |:      |  |: ||    ||
Db    104472 GCCGGCCSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGGCSNNNDNNCCGCBGGCCBGGGC 104531

Qy       306 GGATCACGCGCCGCGGGCCGCA--CACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGG 363
             |   | |  |:   :  ||||   | :|| |||  || ||:   :  ||||:||| : ||
Db    104532 GCGCCGCCGGSNNNDNNCCGCBGGCCBGGGCGCGCCGCCGSNNNDNNCCGCBGGCCBGGG 104591

Qy       364 CTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGC 423
             | |||||  :   :  | |   || :   || ||       :   |||:   |:| ||||
Db    104592 CGCGCCGCCSNNNDNNCCGCBGGCCBGGGCGCGCCGCSNNNDNNCCGCBGGCCBGGGCGC 104651

Qy       424 GCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCA 483
             ||| :       |||:   |: ||||||  :   :    ||:  |   |  ||||:   :
Db    104652 GCCGSNNNDNNCCGCBGGCCBGGGCGCGCCSNNNDNNCCGCBGGCCBGGGCGCGCSNNND 104711

Qy       484 GAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAG 543
             |  |:|           :   :    ||:  | :| ||  :   :   || :|  | |
Db    104712 NNCCGCBGGCCBGGGCGCGSNNNDNNCCGCBGGCCBGGGCGCSNNNDNNCCGCBGGCCBG 104771

Qy       544 CGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAG 603
             | |  :     :|  |:||  :     :  |:   :   :   |   :|
Db    104772 GGCGSNNNDNNCCGCBGGCCBGGGCSNNNDNNCCGCBGGCCBGGGSNNNDNNCCGCBGGC 104831

Qy       604 CTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTT 663
             |:  :   :   |||:    :           | :| | ::   :      :|  |    :
Db    104832 CBGGSNNNDNNCCGCBGGCCBGSNNNDNNCCGCBGGCCBSNNNDNNCCGCBGGCCSNNND 104891

Qy       664 GTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGT 723
             || :  |   ||    |||      ||    |   :  | :   |:  | |   ||
Db    104892 NNCGCBGGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNGCBGGCCBGGGCGCGCCGCC 104951

Qy       724 CACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTG 783
             |   | |  :   :   :|| | |  | |   | |   |  |         :  :| | :|
Db    104952 GGCCGGGCCGSNNNDNNCBGGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNBGGCCBG 105011

Qy       784 ATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCG 843
             ||    ||  || ||   |:   :   || :|| | || ||  | | |||  |    :
Db    105012 GGCGCGCCGCCGGCCGGGCCGSNNNDNNGGCCBGGGCGCGCCGCCGGCCGGGCCGSNNND 105071

Qy       844 GGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGG 884
             | |:| || ||| | |||| | |   :   :   |: ||
Db    105072 NNGCCBGGGCGCGCCGCCGGCCGGGCCGSNNNDNNCCBGGG 105112
```

RESULT 3
ADD26442/c
ID ADD26442 standard; DNA; 11604 BP.
XX
AC ADD26442;
XX
DT 15-JAN-2004 (first entry)
XX
DE Streptomyces clavuligerus 11.6 kb genomic fragment.
XX
KW ds; clavulanic acid synthesis; beta-lactamase inhibitor; clavulanic acid.
XX
OS Streptomyces clavuligerus.
XX
PN US6589775-B1.
XX
PD 08-JUL-2003.
XX
PF 30-AUG-1999; 99US-00385040.
XX
PR 08-OCT-1993; 93US-00134018.
PR 06-DEC-1995; 95US-00567801.
PR 29-JAN-1997; 97US-00790462.
XX
PA (UYAL-) UNIV ALBERTA.
XX
PI Jensen SE, Aidoo KA, Paradkar AS;
XX
DR WPI; 2003-810383/76.
XX
PT New isolated DNA molecule encoding enzymes of clavulanic acid
PT biosynthesis, useful for producing and enhancing production of clavulanic
PT acid in a transformant host.
XX
PS Disclosure; SEQ ID NO 13; 67pp; English.
XX
CC The invention relates to an isolated DNA molecule. The methods and
CC compositions of the present invention are useful for producing and
CC enhancing production of clavulanic acid (beta-lactamase inhibitor) in a
CC clavulanic acid producing host. The present sequence represents a 15kb
CC genomic fragment from Streptomyces clavuligerus.
XX
SQ Sequence 11604 BP; 1649 A; 4295 C; 4038 G; 1622 T; 0 U; 0 Other;

Query Match 9.0%; Score 81.2; DB 10; Length 11604;
Best Local Similarity 45.1%; Pred. No. 0.00037;
Matches 342; Conservative 0; Mismatches 413; Indels 3; Gaps 1;

```
Qy       114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
             || ||    |  || || || ||    |  ||| |    |  |     |  || | ||
Db      1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022
```

```
Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    ||||  |      ||| |||  ||
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               |||  ||| | |   |   | ||| |     ||   || |   |    |||| ||  | |
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||          |||  | || || |  | ||| |||  || ||     ||
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||     || | || || |  |  |||   | ||| | ||
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
               |||          || |  ||  |  || ||  |  |  |  |||||||  | |||
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |       ||||   | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        601 GGAGCTCCACGGCGTACAAGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  |  |||  ||   | | ||     |  |||  |||  |     |
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |        || |  |     |||    |||   |    |   |  | |
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||      | |  | |  ||  |    |||||||    || ||  || |
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCG 868
              ||| ||  ||| || ||   ||  ||  | ||  || |
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCGATGGGG 324
```

RESULT 4
ADG47787/c
ID ADG47787 standard; DNA; 11604 BP.
XX
AC ADG47787;
XX
DT 11-MAR-2004 (first entry)
XX
DE Streptomyces clavuligerus 15 kb gene fragment DNA.
XX
KW Clavulanic acid; bacterial infection; gene therapy; ds.
XX
OS Streptomyces clavuligerus.
XX
PN US2003207411-A1.
XX
PD 06-NOV-2003.
XX
PF 11-JUN-2003; 2003US-00458201.
XX
PR 08-OCT-1993; 93US-00134018.
PR 06-DEC-1995; 95US-00567801.
PR 29-JAN-1997; 97US-00790462.
PR 30-AUG-1999; 99US-00385040.
XX
PA (UYAL-) UNIV ALBERTA.
XX
PI Jensen SE, Aidoo KA, Paradkar AS;
XX
DR WPI; 2003-875866/81.
XX
PT New DNA molecule, useful in producing clavulanic acid in a non-
PT clavulanate-producing host or in enhancing clavulanic acid production in
PT a clavulanate-producing host for preparing a composition for treating
PT bacterial infections.
XX
PS Disclosure; SEQ ID NO 13; 69pp; English.
XX
CC The present invention relates to 15 kb DNA which encodes 10 ORFs. The DNA
CC molecule is useful in producing clavulanic acid in a non-clavulanate-
CC producing host or in enhancing clavulanic acid production in a
CC clavulanate-producing host for preparing a composition for treating
CC bacterial infections. The invention is also useful in gene therapy. The
CC present sequence is Streptomyces clavuligerus 15 kb gene fragment DNA.
XX
SQ Sequence 11604 BP; 1649 A; 4295 C; 4038 G; 1622 T; 0 U; 0 Other;

Query Match 9.0%; Score 81.2; DB 10; Length 11604;
Best Local Similarity 45.1%; Pred. No. 0.00037;
Matches 342; Conservative 0; Mismatches 413; Indels 3; Gaps 1;

```
Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |  || || || ||    | ||| |   |  |     |  || | ||
Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    ||||  |      ||| |||  ||
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               |||  ||| | |   |   | ||| |     ||   || |   |    |||| ||  | |
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||          |||  | || || |  | ||| |||  || ||     ||
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||     || | || || |  |  |||   | ||| | ||
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
               |||          || |  ||  |  || ||  |  |  |  |||||||  | |||
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |       ||||   | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        601 GGAGCTCCACGGCGTACAAGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  |  |||  ||   | | ||     |  |||  |||  |     |
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |        || |  |     |||    |||   |    |   |  | |
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
```

```
                    ||     | |  | |  ||  |    |||||||    || ||  || |
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCG 868
              ||| ||  ||| || ||   ||  ||  | ||  || |
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCGATGGGG 324
```

RESULT 5
AEB57391
ID AEB57391 standard; DNA; 3024 BP.
XX
AC AEB57391;
XX
DT 22-SEP-2005 (first entry)
XX
DE Silk elastin repeat sequence protein copolymer encoding DNA SEQ ID NO:35.
XX
KW repeat sequence protein polymer; cosmetics; dermatological; fungicide;
KW antiseborrheic; gene; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1. .3024
FT /*tag= a
FT /product= "silk elastin repeat sequence protein
FT copolymer"
FT /note= "the sequence encodes amino acids 34 to 1041 of
FT SEQ ID NO:34"
XX
PN US2005142094-A1.
XX
PD 30-JUN-2005.
XX
PF 10-SEP-2004; 2004US-00939036.
XX
PR 12-MAR-2003; 2003US-0454077P.
PR 12-MAR-2004; 2004US-00800179.
XX
PA (KUMA/) KUMAR M.
XX
PI Kumar M;
XX
DR WPI; 2005-561167/57.
DR P-PSDB; AEB57390.
XX
PT Personal care composition useful for protecting skin from environmental
PT damage, has carrier and repeat sequence protein polymer capable of
PT providing cosmetic benefit such as increasing skin fibroblast viability.
XX
PS Example 5; SEQ ID NO 35; 107pp; English.
XX
CC The invention relates to a personal care composition (I) comprising: (a)
CC a repeat sequence protein polymer (RSPP) or its fragment, and a carrier
CC or excipient, where repeat sequence protein polymer provides at least one
CC cosmetic benefit; or (b) a RSPP or its fragment, and one or more
CC compounds chosen from carriers, excipients, liposomes, biological or
CC botanical products, and where repeat sequence protein polymer or its
CC fragment improves skin tone. (I) is useful for protecting skin from
CC environmental damage, where the environmental damage comprises damage due
CC to exposure to radiation (especially ultraviolet light) or damage due to
CC exposure to chemicals or free radicals. (I) is useful as a skin care
CC composition such as moisturizing body washes, body washes, antimicrobial
CC cleansers, skin protective creams, body lotions, facial creams,
CC moisturizing creams, facial cleansing emulsions, surfactant-based facial
CC cleansers, facial exfoliating gels, anti-acne treatments, facial toners,
CC exfoliating creams, facial masks, aftershave balms, composition
CC comprising topically applied over-the counter compositions, anti-fungal
CC treatments, skin protectants, anti-perspirants, or radioprotective such
CC as sunscreen chosen from non-water-resistant sunscreens, very water-
CC resistant sunscreens, and water-in silicone sunscreens. (I) is useful as
CC a makeup composition such as eye gels, high-melting point lipsticks,
CC lipsticks, lip glosses, lip balms, mascaras, eyeliners, pressed powder
CC formulations chosen from loose powders, blushes, eye shadows, and
CC bronzing powders, foundation chosen from water-in-oil foundations, water-
CC in-silicone foundations, oil-in-water foundations, anhydrous makeup
CC sticks and cream-to-powder foundations, and as makeup composition
CC comprising at least one pigment chosen from a mascara such as non-
CC waterproof mascaras, waterproof mascaras, voluminizing mascaras,
CC lengthening mascaras, curling mascaras, anhydrous waterproof mascaras,
CC water-based mascaras, and eyelash or eyebrow treatments. (I) enables
CC preparation of personal care products that have desired characteristics
CC (without chemical modifications of the protein) such as transparent film
CC formation, hydrogel formation, better efficacy and binding to skin, hair,
CC nail and oral surfaces, desired level of hydrophobicity with water
CC solubility, imparting luster, softness, moisture retainment, mechanical
CC properties such as tensile properties, viscoelastic behavior, glass
CC transition temperature, cloud temperature, and decomposition temperature.
CC The present sequence encodes a silk elastin repeat sequence protein
CC copolymer RSPP, which is used in the exemplification of the present
CC invention.
XX
SQ Sequence 3024 BP; 160 A; 1216 C; 1632 G; 16 T; 0 U; 0 Other;

Query Match 8.9%; Score 80.2; DB 14; Length 3024;
Best Local Similarity 47.0%; Pred. No. 0.00062;
Matches 422; Conservative 0; Mismatches 458; Indels 17; Gaps 5;

```
Qy          5 GGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCG 64
              ||| ||   ||||  |  |||  || || ||||| || |  |   | |     |||  |
Db        531 GGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCGCAGGGCCCGGCGGGCCC 590

Qy         65 GGCCGCTGAGGTCGCGCGGGC--TACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCG 122
              || ||  | |   |  | |||    | ||||| |||| || | |||   |||   |   |
Db        591 GGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGG 650

Qy        123 CCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGA 182
              | ||   |      |         | ||  |   ||| |  || |  | ||  ||  |   |
Db        651 CGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCA 710

Qy        183 TGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCA 242
              || |  | || ||    |||  | ||  ||||  |    | ||  | |||| | |  |
Db        711 TGGCCCGGCGGGCCCGAAAGGCGCGCATGGCCCGGCGGGCCCGAAAGGCGCGCATGGCCC 770

Qy        243 CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGAT 302
              |    | |     | || ||   |||  || | |||  |    |   |  |||  |||
Db        771 GGCGGGCCCGAAAGGCGCGCATGGCCCGGCGGGCCCGAAAGGCGCGCAGGGCCCGGCGGG 830

Qy        303 CGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCG 362
              | |||     ||||| | |  | ||    |    || ||| |||    | | ||||||   |
Db        831 CCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCG-CGCAGGGCCCGGCGGGCCCGG 889

Qy        363 GCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCG 422
              ||    || || |  ||  |  | ||| |        |  ||  |  | | || |  |||
Db        890 GCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCG 949

Qy        423 CGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGC 482
              ||| |   | |  |    | ||||||  | ||  |  | ||||   |    ||  ||
Db        950 CGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCC---GGGCGGCGCGC 1006

Qy        483 AGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCA 542
              ||  | |||||    || |||    |  ||        | ||| || |    ||  | ||
Db       1007 AGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCC 1066

Qy        543 GCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAG------AGATAGC 596
                | |  |   ||    |||| ||  | | ||  | ||    ||  |           |
Db       1067 CGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGG 1126

Qy        597 GCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGA 656
              || || ||  |  | | || | | | | |  | | | |                | | |
Db       1127 GCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGG 1186

Qy        657 GCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGG 716
              || ||  |  |  | ||  |||| |  |    ||| || |  | |  | ||| || | |
Db       1187 GCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCG 1246
```

```
Qy        717 TCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGG 776
              || | ||  | |    ||  | ||| | |  ||  || |||||    |||
Db       1247 CGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGG 1306

Qy        777 TGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGA 836
              |||     |  |  |  | |||| ||  |  | |   |||| |||| ||   ||
Db       1307 GCCCG-----GCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGG 1361

Qy        837 TGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCC 893
              |||||  |||    ||||   |||   | || |||   ||| ||  |||||
Db       1362 CCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCC 1418
```

RESULT 6
AAD14503/c
ID AAD14503 standard; DNA; 1722 BP.
XX
AC AAD14503;
XX
DT 01-NOV-2001 (first entry)
XX
DE S. clavuligerus clavulanic acid biosynthesis enzyme DNA #1.
XX
KW Clavulanic acid biosynthesis enzyme; antibiotic; infectious disease;
KW broad spectrum beta-lactamase inhibitor; open reading frame; ORF;
KW pcbC gene; ds.
XX
OS Streptomyces clavuligerus.
XX
FH Key Location/Qualifiers
FT CDS 1. .1722
FT /*tag= a
FT /product= "Clavulanic acid biosynthesis enzyme"
FT /transl_except= (pos:1717. .1719, aa:Ile-Glx)
XX
PN US6232106-B1.
XX
PD 15-MAY-2001.
XX
PF 30-AUG-1999; 99US-00385028.
XX
PR 08-OCT-1993; 93US-00134018.
PR 06-DEC-1995; 95US-00567801.
PR 29-JAN-1997; 97US-00790462.
XX
PA (UYAL-) UNIV ALBERTA.
XX
PI Jensen SE, Aidoo KA, Paradkar AS;
XX
DR WPI; 2001-342772/36.
DR P-PSDB; AAE07907.
XX
PT Novel enzyme required for clavulanic acid biosynthesis which is useful as
PT broad spectrum beta-lactamase inhibitor.
XX
PS Example 6; Col 63-66; 75pp; English.
XX
CC The invention relates to DNA sequences encoding enzymes required for
CC clavulanic acid biosynthesis. Clavulanic acid is a broad spectrum beta-
CC lactamase inhibitor and is an important antibiotic for the treatment of
CC infectious diseases. Also provided in the patent is a 15 Kb genomic DNA
CC fragment downstream to pcbC gene from Streptomyces clavuligerus. The
CC genomic DNA comprises 10 open reading frames (ORFs), eight of which are
CC involved in clavulanic acid biosynthesis. The present sequence is S.
CC clavuligerus ORF2 DNA encoding enzyme involved in clavulanic acid
CC biosynthesis
XX
SQ Sequence 1722 BP; 263 A; 680 C; 511 G; 268 T; 0 U; 0 Other;

Query Match 8.9%; Score 80; DB 4; Length 1722;
Best Local Similarity 45.1%; Pred. No. 0.00071;
Matches 339; Conservative 0; Mismatches 410; Indels 3; Gaps 1;

```
Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||     |  || || || ||    | ||| |   |  |    |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |  |||| ||  |    |  | ||| ||| |    ||||  |      ||| |||  ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |  |   | ||| |     ||   || |   |    |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||           |||   | || || |  | ||| |||  || ||     ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |   || |||   || || | || ||| ||   |||   |||  |      |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||         || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||          ||  |  ||  |   || ||  |  |  | |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |   ||||   | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  |  |||  ||   | | ||   | |||  |||  |      |
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |      |||    |||  |    |  |  | | |
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||      | |  | |  ||  |    |||||||    || ||  || |
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147
```

RESULT 7
ADD26444/c
ID ADD26444 standard; DNA; 1722 BP.
XX
AC ADD26444;
XX
DT 15-JAN-2004 (first entry)
XX
DE Streptomyces clavuligerus ORF2 DNA.
XX
KW ds; clavulanic acid synthesis; beta-lactamase inhibitor; clavulanic acid;
KW gene.
XX
OS Streptomyces clavuligerus.
XX
FH Key Location/Qualifiers
FT CDS 1. .1722
FT /*tag= a
FT /transl_except= (pos:1720. .1722,aa:Glx)

```
XX
PN   US6589775-B1.
XX
PD   08-JUL-2003.
XX
PF   30-AUG-1999;   99US-00385040.
XX
PR   08-OCT-1993;   93US-00134018.
PR   06-DEC-1995;   95US-00567801.
PR   29-JAN-1997;   97US-00790462.
XX
PA   (UYAL-) UNIV ALBERTA.
XX
PI   Jensen SE,  Aidoo KA,  Paradkar AS;
XX
DR   WPI; 2003-810383/76.
DR   P-PSDB; ADD26433.
XX
PT   New isolated DNA molecule encoding enzymes of clavulanic acid
PT   biosynthesis, useful for producing and enhancing production of clavulanic
PT   acid in a transformant host.
XX
PS   Claim 1; SEQ ID NO 15; 67pp; English.
XX
CC   The invention relates to an isolated DNA molecule. The methods and
CC   compositions of the present invention are useful for producing and
CC   enhancing production of clavulanic acid (beta-lactamase inhibitor) in a
CC   clavulanic acid producing host. The present sequence represents DNA
CC   encoding an ORF involved in clavulanic acid synthesis in Streptomyces
CC   clavuligerus.
XX
SQ   Sequence 1722 BP; 263 A; 680 C; 511 G; 268 T; 0 U; 0 Other;

  Query Match             8.9%;  Score 80;  DB 10;  Length 1722;
  Best Local Similarity   45.1%;  Pred. No. 0.00071;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||     |   || || || ||     |  ||| |     |  |      |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    ||||  |       ||| |||   ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |   |  | ||| |     ||   || |   |    |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||          |||   | || || |  | ||| |||  || ||      ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||        || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||   | |  | ||| ||    |  |   | ||   | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||      |  | |||  ||  | | ||  |  |||  |||  |     |
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              || | |        || | |    |||   |||   |   |   |  | | |
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                  ||      | | | | ||  |    |||||||    || ||  || |
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 8
ADG47789/c
ID   ADG47789 standard; DNA; 1722 BP.
XX
AC   ADG47789;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Streptomyces clavuligerus 15 kb gene ORF2 DNA.
XX
KW   Clavulanic acid; bacterial infection; gene therapy; gene;
KW   open reading frame; ORF; ds.
XX
OS   Streptomyces clavuligerus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1722
FT                   /*tag=  a
FT                   /product= "ORF2 protein"
FT                   /transl_except= (pos:1717. .1719, aa:Ile-Glx)
XX
PN   US2003207411-A1.
XX
PD   06-NOV-2003.
XX
PF   11-JUN-2003; 2003US-00458201.
XX
PR   08-OCT-1993;   93US-00134018.
PR   06-DEC-1995;   95US-00567801.
PR   29-JAN-1997;   97US-00790462.
PR   30-AUG-1999;   99US-00385040.
XX
PA   (UYAL-) UNIV ALBERTA.
XX
PI   Jensen SE,  Aidoo KA,  Paradkar AS;
XX
DR   WPI; 2003-875866/81.
DR   P-PSDB; ADG47778.
XX
PT   New DNA molecule, useful in producing clavulanic acid in a non-
PT   clavulanate-producing host or in enhancing clavulanic acid production in
PT   a clavulanate-producing host for preparing a composition for treating
PT   bacterial infections.
XX
PS   Claim 1; SEQ ID NO 15; 69pp; English.
XX
CC   The present invention relates to 15 kb DNA which encodes 10 ORFs. The DNA
CC   molecule is useful in producing clavulanic acid in a non-clavulanate-
CC   producing host or in enhancing clavulanic acid production in a
CC   clavulanate-producing host for preparing a composition for treating
CC   bacterial infections. The invention is also useful in gene therapy. The
CC   present sequence is Streptomyces clavuligerus 15 kb gene ORF DNA.
XX
SQ   Sequence 1722 BP; 263 A; 680 C; 511 G; 268 T; 0 U; 0 Other;

  Query Match             8.9%;  Score 80;  DB 10;  Length 1722;
```

```
  Best Local Similarity   45.1%;  Pred. No. 0.00071;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |      |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||   |    |   | ||| ||| |    ||||  |       ||| |||  ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |   |   | ||| |      ||   || |   |    |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||          |||   | || || |  | ||| |||  || ||      ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||          || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||   | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  |  |||  ||   | | ||   |  |||  |||  |       |
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |      |||     |||   |     |   |  | ||
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||      | |  | |  ||  |    |||||||    || ||  || |
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 9
ACA62926/c
ID   ACA62926 standard; DNA; 1722 BP.
XX
AC   ACA62926;
XX
DT   26-AUG-2003  (first entry)
XX
DE   DNA encoding clavulanic acid synthesis associated protein #2.
XX
KW   Clavulanic acid synthesis; broad-spectrum beta-lactamase inhibitor;
KW   antibiotic; infectious disease; antibiotic yield; antimicrobial;
KW   beta-lactamase inhibitor; gene; ds.
XX
OS   Streptomyces clavuligerus.
XX
PN   US6514735-B1.
XX
PD   04-FEB-2003.
XX
PF   01-DEC-2000; 2000US-00726614.
XX
PR   08-OCT-1993;   93US-00134018.
PR   06-DEC-1995;   95US-00567801.
PR   29-JAN-1997;   97US-00790462.
PR   30-AUG-1999;   99US-00385028.
XX
PA   (UYAL-) UNIV ALBERTA.
XX
PI   Jensen SE,  Aidoo KA,  Paradkar AS;
XX
DR   WPI; 2003-491699/46.
DR   P-PSDB; ABU62215.
XX
PT   New protein/enzyme encoded by a DNA from Streptomyces clavuligerus,
PT   useful in the synthesis of clavulanic acid, which is a broad spectrum
PT   beta-lactamase inhibitor for treating infections, or for increasing
PT   antibiotic yield.
XX
PS   Disclosure; Col 63-66; 76pp; English.
XX
CC   The invention describes an isolated protein, which comprises a sequence
CC   having 409 amino acids fully defined in the specification. The protein is
CC   useful in the synthesis of clavulanic acid, which is a broad-spectrum
CC   beta-lactamase inhibitor and an important antibiotic for treating
CC   infectious diseases. The protein is particularly useful for increasing
CC   antibiotic yield. This sequence encodes a clavulanic acid synthesis
CC   associated protein
XX
SQ   Sequence 1722 BP; 263 A; 680 C; 511 G; 268 T; 0 U; 0 Other;

  Query Match             8.9%;  Score 80;  DB 10;  Length 1722;
  Best Local Similarity   45.1%;  Pred. No. 0.00071;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |      |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||   |    |   | ||| ||| |    ||||  |       ||| |||  ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |   |   | ||| |      ||   || |   |    |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||          |||   | || || |  | ||| |||  || ||      ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||          || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||   | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359
```

```
Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147
```

RESULT 10
AAD14501/c
ID AAD14501 standard; DNA; 11604 BP.
XX
AC AAD14501;
XX
DT 01-NOV-2001 (first entry)
XX
DE Streptomyces clavuligerus genomic DNA fragment.
XX
KW Clavulanic acid biosynthesis enzyme; antibiotic; infectious disease;
KW broad spectrum beta-lactamase inhibitor; ds.
XX
OS Streptomyces clavuligerus.
XX
PN US6232106-B1.
XX
PD 15-MAY-2001.
XX
PF 30-AUG-1999; 99US-00385028.
XX
PR 08-OCT-1993; 93US-00134018.
PR 06-DEC-1995; 95US-00567801.
PR 29-JAN-1997; 97US-00790462.
XX
PA (UYAL-) UNIV ALBERTA.
XX
PI Jensen SE, Aidoo KA, Paradkar AS;
XX
DR WPI; 2001-342772/36.
DR P-PSDB; AAE07912, AAE07913, AAE07914, AAE07915.
XX
PT Novel enzyme required for clavulanic acid biosynthesis which is useful as
PT broad spectrum beta-lactamase inhibitor.
XX
PS Disclosure; Col 51-62; 75pp; English.
XX
CC The invention relates to DNA sequences encoding enzymes required for
CC clavulanic acid biosynthesis. Clavulanic acid is a broad spectrum beta-
CC lactamase inhibitor and is an important antibiotic for the treatment of
CC infectious diseases. Also provided in the patent is a 15 Kb genomic DNA
CC fragment from Streptomyces clavuligerus. The genomic DNA comprises 10
CC open reading frames (ORFs), eight of which are involved in clavulanic
CC acid biosynthesis. The present sequence is a genomic DNA fragment from S.
CC clavuligerus
XX
SQ Sequence 11604 BP; 1651 A; 4290 C; 4042 G; 1621 T; 0 U; 0 Other;

Query Match 8.9%; Score 80; DB 4; Length 11604;
Best Local Similarity 45.1%; Pred. No. 0.00057;
Matches 339; Conservative 0; Mismatches 410; Indels 3; Gaps 1;

```
Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
Db        601 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 330
```

RESULT 11
ACA62924/c
ID ACA62924 standard; DNA; 11604 BP.
XX
AC ACA62924;
XX
DT 26-AUG-2003 (first entry)
XX
DE Clavulanic acid synthesis associated DNA.
XX
KW Clavulanic acid synthesis; broad-spectrum beta-lactamase inhibitor;
KW antibiotic; infectious disease; antibiotic yield; antimicrobial;
KW beta-lactamase inhibitor; ds.
XX
OS Unidentified.
XX
PN US6514735-B1.
XX
PD 04-FEB-2003.
XX
PF 01-DEC-2000; 2000US-00726614.
XX
PR 08-OCT-1993; 93US-00134018.

```
PR   06-DEC-1995;   95US-00567801.
PR   29-JAN-1997;   97US-00790462.
PR   30-AUG-1999;   99US-00385028.
XX
PA   (UYAL-) UNIV ALBERTA.
XX
PI   Jensen SE,  Aidoo KA,  Paradkar AS;
XX
DR   WPI; 2003-491699/46.
DR   P-PSDB; ABU62220, ABU62221, ABU62222, ABU62223.
XX
PT   New protein/enzyme encoded by a DNA from Streptomyces clavuligerus,
PT   useful in the synthesis of clavulanic acid, which is a broad spectrum
PT   beta-lactamase inhibitor for treating infections, or for increasing
PT   antibiotic yield.
XX
PS   Disclosure; Col 51-62; 76pp; English.
XX
CC   The invention describes an isolated protein, which comprises a sequence
CC   having 409 amino acids fully defined in the specification. The protein is
CC   useful in the synthesis of clavulanic acid, which is a broad-spectrum
CC   beta-lactamase inhibitor and an important antibiotic for treating
CC   infectious diseases. The protein is particularly useful for increasing
CC   antibiotic yield. This sequence represents a polynucleotide associated
CC   with Streptomyces clavuligerus 15kb genomic sequence that encodes enzymes
CC   required for clavulanic acid synthesis
XX
SQ   Sequence 11604 BP; 1651 A; 4290 C; 4042 G; 1621 T; 0 U; 0 Other;

  Query Match             8.9%;  Score 80;  DB 10;  Length 11604;
  Best Local Similarity   45.1%;  Pred. No. 0.00057;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    | ||| |    |  |     |  || | ||
Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    ||||  |       ||| |||  ||
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |  |   | ||| |    ||   || |   |    |||| ||  | |
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||         |||   | || || |  | ||| |||  || ||      ||
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  | | |  ||  ||  | |||
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |     |
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||      || | || || |  |  |||   | ||| | ||
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||         || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |     ||||   | |  | ||| ||   |  |   | ||  | | ||| |  || ||
Db        601 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  |  |||  ||   | | ||   |  |||  |||  |      |
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |        || |  |      |||    |||    |     |   |  | ||
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||     |  |  | |  ||  |    |||||||    || ||  || |
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 330


RESULT 12
AAD14499/c
ID   AAD14499 standard; DNA; 15079 BP.
XX
AC   AAD14499;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Streptomyces clavuligerus 15 Kb genomic DNA fragment.
XX
KW   Clavulanic acid biosynthesis enzyme; antibiotic; infectious disease;
KW   broad spectrum beta-lactamase inhibitor; ds.
XX
OS   Streptomyces clavuligerus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(109. .1764)
FT                   /*tag=  a
FT                   /product= "Open reading frame (ORF)1 protein"
FT                   /transl_except= (pos:112. .114, aa:Gly-Glx)
FT   misc_feature    1765. .2215
FT                   /*tag=  b
FT                   /note= "Intergenic region"
FT   CDS             2216. .3937
FT                   /*tag=  c
FT                   /product= "Clavulanic acid biosynthesis enzyme"
FT                   /transl_except= (pos:3932. .3934, aa:Ile-Glx)
FT                   /note= "Corresponds to ORF2"
FT   misc_feature    3938. .3939
FT                   /*tag=  d
FT                   /note= "Intergenic region"
FT   CDS             3940. .5481
FT                   /*tag=  e
FT                   /product= "Clavulanic acid biosynthesis enzyme"
FT                   /transl_except= (pos:5476. .5478, aa:Ala-Glx)
FT                   /note= "Corresponds to ORF3"
FT   misc_feature    5482. .5653
FT                   /*tag=  f
FT                   /note= "Intergenic region"
FT   CDS             5654. .6595
FT                   /*tag=  g
FT                   /product= "Clavulanic acid biosynthesis enzyme, CLA"
FT                   /transl_except= (pos:6590. .6592, aa:Leu-Glx)
FT                   /note= "The CDS corresponds to ORF4 and does not include
FT                   start codon"
FT                   /partial
FT   misc_feature    6596. .6610
FT                   /*tag=  h
FT                   /note= "Intergenic region"
FT   CDS             6611. .7588
FT                   /*tag=  i
FT                   /product= "Clavaminate synthase isozyme (cs2)"
FT                   /transl_except= (pos:7583. .7585, aa:Arg-Glx)
FT                   /note= "Corresponds to ORF5"
FT   misc_feature    7589. .7894
FT                   /*tag=  j
FT                   /note= "Intergenic region"
FT   CDS             7895. .9076
FT                   /*tag=  k
FT                   /product= "Clavulanic acid biosynthesis enzyme"
```

```
FT                   /transl_except= (pos:9071. .9073, aa:Thr-Glx)
FT                   /note= "Corresponds to ORF6"
FT   misc_feature    9077. .9240
FT                   /*tag= l
FT                   /note= "Intergenic region"
FT   CDS             9241. .10908
FT                   /*tag= m
FT                   /product= "Clavulanic acid biosynthesis enzyme"
FT                   /transl_except= (pos:10903. .10905, aa:Lys-Glx)
FT                   /note= "Corresponds to ORF7"
FT   misc_feature    10909. .10997
FT                   /*tag= n
FT                   /note= "Intergenic region"
FT   CDS             complement(10998. .12296)
FT                   /*tag= o
FT                   /product= "Clavulanic acid biosynthesis enzyme"
FT                   /transl_except= (pos:11001. .11003, aa:Gly-Glx)
FT                   /note= "Corresponds to ORF8"
FT   misc_feature    12297. .12621
FT                   /*tag= p
FT                   /note= "Intergenic region"
FT   CDS             complement(12622. .13365)
FT                   /*tag= q
FT                   /product= "Clavulanic acid biosynthesis enzyme"
FT                   /transl_except= (pos:12625. .12627, aa:Val-Glx)
FT                   /note= "Corresponds to ORF9"
FT   misc_feature    13366. .13768
FT                   /*tag= r
FT                   /note= "Intergenic region"
FT   CDS             13769. .14995
FT                   /*tag= s
FT                   /product= "ORF10 protein"
FT                   /transl_except= (pos:14990. .14992, aa:Trp-Glx)
XX
PN   US6232106-B1.
XX
PD   15-MAY-2001.
XX
PF   30-AUG-1999;   99US-00385028.
XX
PR   08-OCT-1993;   93US-00134018.
PR   06-DEC-1995;   95US-00567801.
PR   29-JAN-1997;   97US-00790462.
XX
PA   (UYAL-) UNIV ALBERTA.
XX
PI   Jensen SE,  Aidoo KA,  Paradkar AS;
XX
DR   WPI; 2001-342772/36.
DR   P-PSDB; AAE07906, AAE07907, AAE07908, AAE07909, AAE07910, AAE07911,
DR   AAE07912, AAE07913, AAE07914, AAE07915.
XX
PT   Novel enzyme required for clavulanic acid biosynthesis which is useful as
PT   broad spectrum beta-lactamase inhibitor.
XX
PS   Example 1; Fig 2; 75pp; English.
XX
CC   The present invention relates to DNA sequences encoding enzymes required
CC   for clavulanic acid biosynthesis. Clavulanic acid is a broad spectrum
CC   beta-lactamase inhibitor and is an important antibiotic for the treatment
CC   of infectious diseases. The present sequence is Streptomyces clavuligerus
CC   15 Kb genomic DNA fragment comprising 10 open reading frames (ORFs),
CC   eight of which are involved in clavulanic acid biosynthesis
XX
SQ   Sequence 15079 BP; 2105 A; 5475 C; 5331 G; 2168 T; 0 U; 0 Other;

  Query Match             8.9%;  Score 80;  DB 4;  Length 15079;
  Best Local Similarity   45.1%;  Pred. No. 0.00056;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |  || || || ||    |  ||| |   |  |     |  || | ||
Db       3113 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3054

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |  |||| ||  |   |  | ||| ||| |    ||||  |      ||| ||| ||
Db       3053 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 2994

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               ||| ||| | |  |  | ||| |      ||   ||  |   |    |||| ||  | |
Db       2993 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2934

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||          |||  | || || |  | ||| ||| || ||      ||
Db       2933 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2874

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db       2873 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2814

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |   || |||   || || | || ||| ||   |||   ||| |     |
Db       2813 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2754

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  ||| | ||| | ||
Db       2753 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2694

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||          || |  ||  |   || ||  |  |  |  |||||||  | |||
Db       2693 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2634

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |  ||||  | | |  ||| ||   |  |  | ||  | | ||| |  || ||
Db       2633 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2574

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||      |  |  |||  ||   | | ||   |  |||  |||  |     |
Db       2573 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2514

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |      |||    |||   |   |  |  | | |
Db       2513 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2454

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||     | |  | |  ||  |   |||||||    || ||  || |
Db       2453 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2394

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||    ||  ||  | ||
Db       2393 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2362


RESULT 13
ACA62922/c
ID   ACA62922 standard; DNA; 15079 BP.
XX
AC   ACA62922;
XX
DT   26-AUG-2003  (first entry)
XX
DE   Streptomyces clavuligerus 15kb genomic DNA sequence.
XX
KW   Clavulanic acid synthesis; broad-spectrum beta-lactamase inhibitor;
KW   antibiotic; infectious disease; antibiotic yield; antimicrobial;
KW   beta-lactamase inhibitor; gene; ds.
XX
OS   Streptomyces clavuligerus.
XX
PN   US6514735-B1.
XX
PD   04-FEB-2003.
XX
PF   01-DEC-2000; 2000US-00726614.
XX
PR   08-OCT-1993;   93US-00134018.
```

PR 06-DEC-1995; 95US-00567801.
PR 29-JAN-1997; 97US-00790462.
PR 30-AUG-1999; 99US-00385028.
XX
PA (UYAL-) UNIV ALBERTA.
XX
PI Jensen SE, Aidoo KA, Paradkar AS;
XX
DR WPI; 2003-491699/46.
DR P-PSDB; ABU62214, ABU62215, ABU62216, ABU62217, ABU62218, ABU62219,
DR ABU62220, ABU62221, ABU62222, ABU62223.
XX
PT New protein/enzyme encoded by a DNA from Streptomyces clavuligerus,
PT useful in the synthesis of clavulanic acid, which is a broad spectrum
PT beta-lactamase inhibitor for treating infections, or for increasing
PT antibiotic yield.
XX
PS Example 5; Fig 2A-J; 76pp; English.
XX
CC The invention describes an isolated protein, which comprises a sequence
CC having 409 amino acids fully defined in the specification. The protein is
CC useful in the synthesis of clavulanic acid, which is a broad-spectrum
CC beta-lactamase inhibitor and an important antibiotic for treating
CC infectious diseases. The protein is particularly useful for increasing
CC antibiotic yield. This sequence represents a Streptomyces clavuligerus
CC 15kb genomic sequence that encodes enzymes required for clavulanic acid
CC synthesis
XX
SQ Sequence 15079 BP; 2105 A; 5475 C; 5331 G; 2168 T; 0 U; 0 Other;

Query Match 8.9%; Score 80; DB 10; Length 15079;
Best Local Similarity 45.1%; Pred. No. 0.00056;
Matches 339; Conservative 0; Mismatches 410; Indels 3; Gaps 1;

```
Qy       114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
Db      3113 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3054

Qy       174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
Db      3053 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 2994

Qy       234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
Db      2993 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2934

Qy       291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
Db      2933 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2874

Qy       351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
Db      2873 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2814

Qy       411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
Db      2813 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2754

Qy       471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
Db      2753 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2694

Qy       531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
Db      2693 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2634

Qy       591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
Db      2633 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2574

Qy       651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
Db      2573 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2514

Qy       711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
Db      2513 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2454

Qy       771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
Db      2453 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2394

Qy       831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
Db      2393 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2362
```

RESULT 14
ADD26455/c
ID ADD26455 standard; DNA; 15120 BP.
XX
AC ADD26455;
XX
DT 15-JAN-2004 (first entry)
XX
DE Streptomyces clavuligerus 15 kb genomic fragment.
XX
KW ds; clavulanic acid synthesis; beta-lactamase inhibitor; clavulanic acid.
XX
OS Streptomyces clavuligerus.
XX
FH Key Location/Qualifiers
FT misc_feature 1765. .5694
FT /*tag= a
FT /note= "Specifically claimed in claim 27"
FT misc_feature 2257. .13406
FT /*tag= c
FT /note= "Specifically claimed in claim 29"
FT misc_feature 2257. .5522
FT /*tag= b
FT /note= "Specifically claimed in claim 25"
FT misc_feature 7630. .13809
FT /*tag= d
FT /note= "Specifically claimed in claim 28"
FT misc_feature 7936. .13406
FT /*tag= e
FT /note= "Specifically claimed in claim 26"
XX
PN US6589775-B1.
XX
PD 08-JUL-2003.
XX
PF 30-AUG-1999; 99US-00385040.
XX
PR 08-OCT-1993; 93US-00134018.
PR 06-DEC-1995; 95US-00567801.
PR 29-JAN-1997; 97US-00790462.
XX
PA (UYAL-) UNIV ALBERTA.
XX
PI Jensen SE, Aidoo KA, Paradkar AS;
XX
DR WPI; 2003-810383/76.
XX
PT New isolated DNA molecule encoding enzymes of clavulanic acid
PT biosynthesis, useful for producing and enhancing production of clavulanic
PT acid in a transformant host.
XX
PS Example 5; SEQ ID NO 1; 67pp; English.
XX
CC The invention relates to an isolated DNA molecule. The methods and
CC compositions of the present invention are useful for producing and
CC enhancing production of clavulanic acid (beta-lactamase inhibitor) in a
CC clavulanic acid producing host. The present sequence represents a 15kb
CC genomic fragment from Streptomyces clavuligerus.
XX

```
SQ   Sequence 15120 BP; 2116 A; 5484 C; 5343 G; 2177 T; 0 U; 0 Other;

  Query Match             8.9%;  Score 80;  DB 10;  Length 15120;
  Best Local Similarity   45.1%;  Pred. No. 0.00056;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |     |  || | ||
Db       3154 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3095

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    ||||  |       ||| |||  ||
Db       3094 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 3035

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |  |   | ||| |     ||   || |   |    |||| ||  | |
Db       3034 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2975

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||          |||   | || || |  | ||| |||  || ||     ||
Db       2974 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2915

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db       2914 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2855

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |   || |||   || || | || ||| ||   |||   ||| |     |
Db       2854 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2795

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db       2794 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2735

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||         || |  ||  |   || ||  |  |  | |||||||  | |||
Db       2734 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2675

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |    ||||  | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db       2674 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2615

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
                ||      | |  ||| ||   | | ||   | ||| ||| |       |
Db       2614 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2555

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |     |||   |||   |     |   |  | ||
Db       2554 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2495

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||     | |  | |  ||  |    |||||||    || ||  || |
Db       2494 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2435

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db       2434 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2403


RESULT 15
ADG47775/c
ID   ADG47775 standard; DNA; 15120 BP.
XX
AC   ADG47775;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Streptomyces clavuligerus 15 kb gene.
XX
KW   Clavulanic acid; bacterial infection; gene therapy; gene; ds.
XX
OS   Streptomyces clavuligerus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(109. .1764)
FT                   /*tag= a
FT                   /product= "ORF1 protein"
FT                   /transl_except= (pos:112. .114, aa:Gly-Glx)
FT   CDS             2257. .3978
FT                   /*tag= b
FT                   /product= "ORF2 protein"
FT                   /transl_except= (pos:3973. .3975, aa:Ile-Glx)
FT   CDS             3981. .5522
FT                   /*tag= c
FT                   /product= "ORF3 protein"
FT                   /transl_except= (pos:5517. .5519, aa:Ala-Glx)
FT   CDS             5695. .6636
FT                   /*tag= d
FT                   /product= "ORF4 protein"
FT                   /transl_except= (pos:6631. .6633, aa:Ala-Glx)
FT                   /note= "No start codon"
FT                   /partial
FT   CDS             6652. .7629
FT                   /*tag= e
FT                   /product= "ORF5 protein"
FT                   /transl_except= (pos:7624. .7626, aa:Arg-Glx)
FT   CDS             7936. .9117
FT                   /*tag= f
FT                   /product= "ORF6 protein"
FT                   /transl_except= (pos:9112. .9114, aa:Thr-Glx)
FT   CDS             9282. .10949
FT                   /*tag= g
FT                   /product= "ORF7 protein"
FT                   /transl_except= (pos:10944. .10946, aa:Lys-Glx)
FT   CDS             complement(11039. .12337)
FT                   /*tag= h
FT                   /product= "ORF8 protein"
FT                   /transl_except= (pos:11042. .11044, aa:Gly-Glx)
FT                   /transl_except= (pos:11534. .11536, aa:Ala)
FT   CDS             complement(12663. .13406)
FT                   /*tag= i
FT                   /product= "ORF9 protein"
FT                   /transl_except= (pos:12668. .12666, aa:Val-Glx)
FT   CDS             13810. .15036
FT                   /*tag= j
FT                   /product= "ORF10 protein"
FT                   /transl_except= (pos:15031. .15033, aa:Trp-Glx)
XX
PN   US2003207411-A1.
XX
PD   06-NOV-2003.
XX
PF   11-JUN-2003; 2003US-00458201.
XX
PR   08-OCT-1993;   93US-00134018.
PR   06-DEC-1995;   95US-00567801.
PR   29-JAN-1997;   97US-00790462.
PR   30-AUG-1999;   99US-00385040.
XX
PA   (UYAL-) UNIV ALBERTA.
XX
PI   Jensen SE,  Aidoo KA,  Paradkar AS;
XX
DR   WPI; 2003-875866/81.
DR   P-PSDB; ADG47777, ADG47778, ADG47779, ADG47780, ADG47781, ADG47782,
DR   ADG47783, ADG47784, ADG47785, ADG47786.
XX
PT   New DNA molecule, useful in producing clavulanic acid in a non-
PT   clavulanate-producing host or in enhancing clavulanic acid production in
PT   a clavulanate-producing host for preparing a composition for treating
PT   bacterial infections.
XX
PS   Claim 31; SEQ ID NO 1; 69pp; English.
XX
CC   The present invention relates to 15 kb DNA which encodes 10 ORFs. The DNA
CC   molecule is useful in producing clavulanic acid in a non-clavulanate-
```

CC producing host or in enhancing clavulanic acid production in a
CC clavulanate-producing host for preparing a composition for treating
CC bacterial infections. The invention is also useful in gene therapy. The
CC present sequence is Streptomyces clavuligerus 15 kb gene.
XX
SQ Sequence 15120 BP; 2116 A; 5484 C; 5343 G; 2177 T; 0 U; 0 Other;

Query Match 8.9%; Score 80; DB 10; Length 15120;
Best Local Similarity 45.1%; Pred. No. 0.00056;
Matches 339; Conservative 0; Mismatches 410; Indels 3; Gaps 1;

```
Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |     |  || | ||
Db       3154 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3095

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    |||| |       ||| |||  ||
Db       3094 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 3035

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               |||  ||| | |  |   | ||| |      ||   || |   |    |||| ||  | |
Db       3034 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2975

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||           |||   | || || |  | ||| |||  || ||      ||
Db       2974 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2915

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   |||||  |  |  | |  ||  ||  | |||
Db       2914 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2855

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |     |
Db       2854 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2795

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db       2794 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2735

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||       || |  ||  |   || ||  |  |  |  |||||||  | |||
Db       2734 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2675

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |    ||||   | |  | ||| ||   |  |   | ||  | | ||| |  || ||
Db       2674 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2615

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  |  |||  ||   | | ||   |  |||  |||  |       |
Db       2614 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2555

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||   | |      || |  |     |||    |||   |   |   |  | |
Db       2554 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2495

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||      | |  | |  ||  |    |||||||    || ||  || |
Db       2494 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2435

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db       2434 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2403
```

Search completed: April 27, 2006, 08:00:17
Job time : 438.473 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 4992.51 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_76000_76900
Perfect score: 901
Sequence: 1 ggcgggggccttccggctgg.....cgggcgcggcccacgcgtcg 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	901	100.0	84868	1	DQ013294	DQ013294 Polyangiu
	2	92	10.2	110000	15	AP008210_184	Continuation (185
	3	92	10.2	133413	15	OSJN00130	AL607003 Oryza sat
	4	92	10.2	148353	15	OSJN00263	AL731623 Oryza sat
c	5	85.6	9.5	1784	1	AF411573S3	AF411575 Actinomad
c	6	83.2	9.2	159997	14	CR352334	CR352334 Danio rer
c	7	82.4	9.1	1094	10	PM7G11B	AL685196 Penicilli
c	8	82.4	9.1	110000	1	BA000030_80	Continuation (81 o
c	9	81.2	9.0	11604	6	AR352525	AR352525 Sequence
c	10	80.8	9.0	110000	1	BX571966_01	Continuation (2 of
	11	80.2	8.9	156789	13	AF533768	AF533768 Cercopith
c	12	80.2	8.9	156789	13	AF533768	AF533768 Cercopith
c	13	80	8.9	1722	6	AR151706	AR151706 Sequence
c	14	80	8.9	1722	6	AR352527	AR352527 Sequence
c	15	80	8.9	11604	6	AR151704	AR151704 Sequence
c	16	80	8.9	15079	6	AR151702	AR151702 Sequence
c	17	80	8.9	15120	1	SCU87786	U87786 Streptomyce
c	18	80	8.9	15120	6	AR352523	AR352523 Sequence
c	19	80	8.9	29870	6	AX763606	AX763606 Sequence
c	20	80	8.9	292200	1	SCO939129	AL939129 Streptomy
c	21	80	8.9	299050	1	SCO939119	AL939119 Streptomy
	22	79.8	8.9	1393	10	PM11H12G	AL684264 Penicilli
c	23	79.6	8.8	33451	1	AY541063	AY541063 Nocardia
c	24	79.4	8.8	142014	14	CR407553	CR407553 Danio rer
	25	79	8.8	150715	13	AY714813	AY714813 Cercopith
c	26	79	8.8	150715	13	AY714813	AY714813 Cercopith
c	27	78.8	8.7	1900	1	SCL238211	AJ238211 Streptomy
	28	78.8	8.7	110000	1	CP000011_12	Continuation (13 o
c	29	78.6	8.7	5317	13	SHK010303	AJ010303 Suid herp
c	30	78.4	8.7	207592	14	CR392332	CR392332 Danio rer
c	31	78	8.7	48000	1	POL505006	AJ505006 Sorangium
c	32	78	8.7	82746	1	AF453501	AF453501 Actinosyn
	33	78	8.7	276800	1	SCO939115	AL939115 Streptomy
	34	77.8	8.6	82746	1	AF453501	AF453501 Actinosyn
	35	77.8	8.6	110000	1	CP000010_12	Continuation (13 o
	36	77.8	8.6	277000	1	SCO939109	AL939109 Streptomy
c	37	77.8	8.6	302007	1	SCO939132	AL939132 Streptomy
	38	77.6	8.6	3756	13	AY318876	AY318876 Pseudorab
c	39	77.4	8.6	1431	6	AX535310	AX535310 Sequence
c	40	77.4	8.6	1579	13	SH1GGIII2	D49436 Pseudorabie
c	41	77.4	8.6	50000	6	AX535264	AX535264 Sequence
	42	77.4	8.6	50000	6	AX535266	AX535266 Sequence
c	43	77.4	8.6	59816	1	AF333038	AF333038 Streptomy
	44	77.4	8.6	136753	14	AC149968	AC149968 Strongylo
c	45	77.2	8.6	77457	1	AF210249	AF210249 Streptomy

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRAREAEREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPFFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

```
                GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
     CDS        complement(1992. .3440)
                /note="ORFu2; similar to Pseudomonas putida KT2440 PP4696;
                putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32957.1"
                /db_xref="GI:63099934"
                /translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
                APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
                QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
                KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
                EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
                VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
                ELTETEIARRLGISRKALWERRQRLGIPRPRS"
     CDS        3587. .5548
                /note="ORFu3; ORFu3p; similar to Leptospira interrogans
                AAN50474"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32958.1"
                /db_xref="GI:63099935"
                /translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
                VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
                LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
                YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
                SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
                IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
                VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
                AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
                GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
                AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
                SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
                VVAGFGATGPLRVMETLPEQYLFLTRVRP"
     CDS        complement(5702. .7231)
                /note="ORFu4; ORFu4p; similar to Burkholderia fungorum
                hypothetical protein in ZP_00033665"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32959.1"
                /db_xref="GI:63099936"
                /translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
                AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
                PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
                VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
                MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
                QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPFSANQRPASTLAQWRVGLEHP
                GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
                GRVITVHFHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
                ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
                P"
     CDS        complement(7561. .8931)
                /note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440
                PP4696; putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32960.1"
                /db_xref="GI:63099937"
                /translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
                APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
                AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
                VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
                VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
                PPPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
                IPRPRS"
     CDS        complement(8928. .11939)
                /note="ORF1; ORF1p; similar to Pseudomonas putida KT2440
                HisK (PP4695) in GenBank Accession Number AAN70268"
                /codon_start=1
                /transl_table=11
                /product="putative sensory box histidine kinase"
                /protein_id="AAY32961.1"
                /db_xref="GI:63099938"
                /translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
                YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
                AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
                FTAFFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
                HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
                FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
                VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
                LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
                IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPAREEREAAR
                LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
                LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
                RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
                LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
                GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
                DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
                DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
                IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
                LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
     CDS        complement(12405. .13379)
                /note="ORF2; ORF2p; similar to family of known or putative
                phosphotransferases, including macrolide
                2-phosphotransferases; similar to Bacillus subtilis YcbJ,
                Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273
                MphB and Staphylococcus aureus plasmid pSR1 MphBM"
                /codon_start=1
                /transl_table=11
                /product="putative macrolide kinase"
                /protein_id="AAY32962.1"
                /db_xref="GI:63099939"
                /translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
                DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
                VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
                IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
                DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
                AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
     CDS        13368. .14663
                /note="ORF3; ORF3p"
                /codon_start=1
                /transl_table=11
                /product="putative Ser/Thr protein kinase"
                /protein_id="AAY32963.1"
                /db_xref="GI:63099940"
                /translation="MFGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
                ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
                RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
                GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
                AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
                DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
                TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
                LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
     gene       15731. .33766
                /gene="dszA"
     CDS        15731. .33766
                /gene="dszA"
                /note="PKS (polyketide synthase) subunit modules 1-4a"
                /codon_start=1
                /transl_table=11
                /product="DszA"
                /protein_id="AAY32964.1"
                /db_xref="GI:63099941"
                /translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
                PRAHYSPDFNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
                IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT
```

```
                     LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                     LSPDGLCKTFDADANGYVPGDGVGVLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                     ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                     RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                     VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                     RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                     DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                     RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                     RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                     ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                     RAQFTFKKFLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                     LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                     PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                     PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                     EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE

  Query Match             100.0%;  Score 901;  DB 1;  Length 84868;
  Best Local Similarity   100.0%;  Pred. No. 3e-70;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83574 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 83633

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83634 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 83693

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83694 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 83753

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83754 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 83813

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83814 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 83873

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83874 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 83933

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83934 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 83993

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83994 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 84053

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84054 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 84113

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84114 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 84173

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84174 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 84233

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84234 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGC 84293

Qy        721 GGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84294 GGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCC 84353

Qy        781 GTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84354 GTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTC 84413

Qy        841 GCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84414 GCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTC 84473

Qy        901 G 901
              |
Db      84474 G 84474


RESULT 2
AP008210_184
WPCOMMENT
  Sequence split into 355 fragments  LOCUS AP008210 Accession AP008210
     Fragment Name      Begin      End
     AP008210_000           1   110000
     AP008210_001      100001   210000
     AP008210_002      200001   310000
     AP008210_003      300001   410000
     AP008210_004      400001   510000
     AP008210_005      500001   610000
     AP008210_006      600001   710000
     AP008210_007      700001   810000
     AP008210_008      800001   910000
     AP008210_009      900001  1010000
     AP008210_010     1000001  1110000
     AP008210_011     1100001  1210000
     AP008210_012     1200001  1310000
     AP008210_013     1300001  1410000
     AP008210_014     1400001  1510000
     AP008210_015     1500001  1610000
     AP008210_016     1600001  1710000
     AP008210_017     1700001  1810000
     AP008210_018     1800001  1910000
     AP008210_019     1900001  2010000
     AP008210_020     2000001  2110000
     AP008210_021     2100001  2210000
     AP008210_022     2200001  2310000
     AP008210_023     2300001  2410000
     AP008210_024     2400001  2510000
     AP008210_025     2500001  2610000
     AP008210_026     2600001  2710000
     AP008210_027     2700001  2810000
     AP008210_028     2800001  2910000
     AP008210_029     2900001  3010000
     AP008210_030     3000001  3110000
     AP008210_031     3100001  3210000
     AP008210_032     3200001  3310000
     AP008210_033     3300001  3410000
     AP008210_034     3400001  3510000
     AP008210_035     3500001  3610000
     AP008210_036     3600001  3710000
     AP008210_037     3700001  3810000
     AP008210_038     3800001  3910000
     AP008210_039     3900001  4010000
     AP008210_040     4000001  4110000
     AP008210_041     4100001  4210000
     AP008210_042     4200001  4310000
     AP008210_043     4300001  4410000
     AP008210_044     4400001  4510000
     AP008210_045     4500001  4610000
     AP008210_046     4600001  4710000
     AP008210_047     4700001  4810000
     AP008210_048     4800001  4910000
     AP008210_049     4900001  5010000
     AP008210_050     5000001  5110000
     AP008210_051     5100001  5210000
     AP008210_052     5200001  5310000
     AP008210_053     5300001  5410000
     AP008210_054     5400001  5510000
     AP008210_055     5500001  5610000
```

AP008210_056	5600001	5710000
AP008210_057	5700001	5810000
AP008210_058	5800001	5910000
AP008210_059	5900001	6010000
AP008210_060	6000001	6110000
AP008210_061	6100001	6210000
AP008210_062	6200001	6310000
AP008210_063	6300001	6410000
AP008210_064	6400001	6510000
AP008210_065	6500001	6610000
AP008210_066	6600001	6710000
AP008210_067	6700001	6810000
AP008210_068	6800001	6910000
AP008210_069	6900001	7010000
AP008210_070	7000001	7110000
AP008210_071	7100001	7210000
AP008210_072	7200001	7310000
AP008210_073	7300001	7410000
AP008210_074	7400001	7510000
AP008210_075	7500001	7610000
AP008210_076	7600001	7710000
AP008210_077	7700001	7810000
AP008210_078	7800001	7910000
AP008210_079	7900001	8010000
AP008210_080	8000001	8110000
AP008210_081	8100001	8210000
AP008210_082	8200001	8310000
AP008210_083	8300001	8410000
AP008210_084	8400001	8510000
AP008210_085	8500001	8610000
AP008210_086	8600001	8710000
AP008210_087	8700001	8810000
AP008210_088	8800001	8910000
AP008210_089	8900001	9010000
AP008210_090	9000001	9110000
AP008210_091	9100001	9210000
AP008210_092	9200001	9310000
AP008210_093	9300001	9410000
AP008210_094	9400001	9510000
AP008210_095	9500001	9610000
AP008210_096	9600001	9710000
AP008210_097	9700001	9810000
AP008210_098	9800001	9910000
AP008210_099	9900001	10010000
AP008210_100	10000001	10110000
AP008210_101	10100001	10210000
AP008210_102	10200001	10310000
AP008210_103	10300001	10410000
AP008210_104	10400001	10510000
AP008210_105	10500001	10610000
AP008210_106	10600001	10710000
AP008210_107	10700001	10810000
AP008210_108	10800001	10910000
AP008210_109	10900001	11010000
AP008210_110	11000001	11110000
AP008210_111	11100001	11210000
AP008210_112	11200001	11310000
AP008210_113	11300001	11410000
AP008210_114	11400001	11510000
AP008210_115	11500001	11610000
AP008210_116	11600001	11710000
AP008210_117	11700001	11810000
AP008210_118	11800001	11910000
AP008210_119	11900001	12010000
AP008210_120	12000001	12110000
AP008210_121	12100001	12210000
AP008210_122	12200001	12310000
AP008210_123	12300001	12410000
AP008210_124	12400001	12510000
AP008210_125	12500001	12610000
AP008210_126	12600001	12710000
AP008210_127	12700001	12810000
AP008210_128	12800001	12910000
AP008210_129	12900001	13010000
AP008210_130	13000001	13110000
AP008210_131	13100001	13210000
AP008210_132	13200001	13310000
AP008210_133	13300001	13410000
AP008210_134	13400001	13510000
AP008210_135	13500001	13610000
AP008210_136	13600001	13710000
AP008210_137	13700001	13810000
AP008210_138	13800001	13910000
AP008210_139	13900001	14010000
AP008210_140	14000001	14110000
AP008210_141	14100001	14210000
AP008210_142	14200001	14310000
AP008210_143	14300001	14410000
AP008210_144	14400001	14510000
AP008210_145	14500001	14610000
AP008210_146	14600001	14710000
AP008210_147	14700001	14810000
AP008210_148	14800001	14910000
AP008210_149	14900001	15010000
AP008210_150	15000001	15110000
AP008210_151	15100001	15210000
AP008210_152	15200001	15310000
AP008210_153	15300001	15410000
AP008210_154	15400001	15510000
AP008210_155	15500001	15610000
AP008210_156	15600001	15710000
AP008210_157	15700001	15810000
AP008210_158	15800001	15910000
AP008210_159	15900001	16010000
AP008210_160	16000001	16110000
AP008210_161	16100001	16210000
AP008210_162	16200001	16310000
AP008210_163	16300001	16410000
AP008210_164	16400001	16510000
AP008210_165	16500001	16610000
AP008210_166	16600001	16710000
AP008210_167	16700001	16810000
AP008210_168	16800001	16910000
AP008210_169	16900001	17010000
AP008210_170	17000001	17110000
AP008210_171	17100001	17210000
AP008210_172	17200001	17310000
AP008210_173	17300001	17410000
AP008210_174	17400001	17510000
AP008210_175	17500001	17610000
AP008210_176	17600001	17710000
AP008210_177	17700001	17810000
AP008210_178	17800001	17910000
AP008210_179	17900001	18010000
AP008210_180	18000001	18110000
AP008210_181	18100001	18210000
AP008210_182	18200001	18310000
AP008210_183	18300001	18410000
AP008210_184	18400001	18510000
AP008210_185	18500001	18610000
AP008210_186	18600001	18710000
AP008210_187	18700001	18810000
AP008210_188	18800001	18910000
AP008210_189	18900001	19010000
AP008210_190	19000001	19110000
AP008210_191	19100001	19210000
AP008210_192	19200001	19310000
AP008210_193	19300001	19410000
AP008210_194	19400001	19510000
AP008210_195	19500001	19610000
AP008210_196	19600001	19710000

Query Match 10.2%; Score 92; DB 15; Length 110000;
Best Local Similarity 45.2%; Pred. No. 2.4;
Matches 405; Conservative 0; Mismatches 470; Indels 21; Gaps 1;

```
Qy          3 CGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCG 62
              | | || | |||   |||    |   |   |||    |  || ||  |        ||| |
Db      30278 CCGAGGTCATCCACTTGGGCTCGCCGGTGAAGCGGACGAGGAGCTCCGCCATCGCCGGGG 30337

Qy         63 CGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCG 122
              ||   |   | | | | | ||  | |     | ||||  | |   | || ||||||
Db      30338 CGTCGCCGAGGAGGCCGGAGAGCCTCCAGTTCCGGACGAGCACCAGGCGGTCGTCGAGCG 30397

Qy        123 CCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGA 182
              |   | |||| ||  ||| ||  | | |  | |||| ||   |  || ||  | |
Db      30398 CGTAGGTGAGGCTGAGGCGGGCGAGGCGCGGGGCGCCGCCGAAGGCCAGCGCGACGGGGC 30457

Qy        183 TGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCCGCAGATCGACGCGGACGAGCA 242
              | |   | ||| | |     |||| | | || | ||||| ||||||  |  ||
Db      30458 CGTCCACGCACGCCACGCGCTCCAGCGCCGGCGCCGCGCGCAGGTCGACGGTGTGGAAGC 30517

Qy        243 CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGAT 302
              ||    || | ||||||||||         |   | ||     ||||  |  | ||   |
Db      30518 CGCAGCCGTCCACGACGAGCTCGCGGAGCTCCGGCATCCCGTCGACCACCAGCGCGGCGT 30577

Qy        303 CGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCG 362
              |    | |||  |  ||| |     |  ||||  |   | |     ||||    ||
Db      30578 ACCACGGCGCGCCGCAGCGGCGGAGGTGGAGCGCCCTGAGCCGCGGGCACGCGTCGAGCA 30637

Qy        363 GCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCG 422
              || || ||||| ||||  |  ||     |  | ||  ||||
Db      30638 CCTGGCGGTAGAAGGCGAGGGGCGTCCGCGGGTGCGTGTCC------------------- 30678

Qy        423 CGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGC 482
              |  | |  ||  | ||  ||  |||| |||| |||                  |  ||
Db      30679 --TGCAGGACGAGCTTGGTGAGCGCGGCGTAGTGCCGGAGCTCGGACGGCGGCGGGAGGA 30736

Qy        483 AGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCA 542
              || | ||       ||||  ||  || |    || || || |  || ||  || |||
Db      30737 AGCAGTTGGACACGGCGAGGCTGCGGATGCGCGACCGGAGCGACTCCGCGCTGAGGCGGT 30796

Qy        543 GCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGA 602
              |   |   |      |||||  | || || |   ||||| |||    | | |   |||
Db      30797 GGTCGGGGAAGCAGAAGGCGTAGCAGGGGCCCTGCTCGACCAGCGCCGGCTTGACGACGG 30856

Qy        603 GCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCT 662
                 ||    | |  |  ||  | | | |   |||| ||   |         |
Db      30857 TGATGTCGAGGTCCTCGACGCCCCACGCGCCGACCGCCATGGCGATCAGGTCCTCGACGA 30916

Qy        663 TGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGG 722
              | |  |   |  | |||  | |  |||   |  |  | |  |||  | |  |||
Db      30917 AGCCCGTGCTGCGCGCCGCGAAGAACTCGAGCCGGAGACTCCTGGCGCGGCGCGGGGCGA 30976

Qy        723 TCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGT 782
              |||  |  ||  | ||||     | |  |   | |  ||| |     | | || || |
Db      30977 CGGCGACGCGCGGCTCCATGAAGGCGCTGACGCAGTCGGCGAACCCGCGGATGGCGCGGC 31036

Qy        783 GATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGC 842
              | |||    |  ||   ||      |||||| |  ||| || |||  | |||  |  ||
Db      31037 GCTCGCACCGGCCGATCGCCGCGTCGAGCTTCCTGGCGTCCGCGCCGGCGGCGTCGCCGT 31096

Qy        843 GGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCG 898
              ||| |||||    |  |  |  |    ||| | | |   || |  ||    || ||
Db      31097 GGGCGCGGACGCAGGCGTGGTAGCGCGGCGGGAGCACGTCGCGGACGCTGAACACG 31152
```

RESULT 3
OSJN00130

```
LOCUS       OSJN00130             133413 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBb0006N15,
            complete sequence.
ACCESSION   AL607003
VERSION     AL607003.3  GI:32489074
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Feng,Q., Zhang,Y., Hao,P., Wang,S., Fu,G., Huang,Y., Li,Y., Zhu,J.,
            Liu,Y., Hu,X., Jia,P., Zhang,Y., Zhao,Q., Ying,K., Yu,S., Tang,Y.,
            Weng,Q., Zhang,L., Lu,Y., Mu,J., Lu,Y., Zhang,L.S., Yu,Z., Fan,D.,
            Liu,X., Lu,T., Li,C., Wu,Y., Sun,T., Lei,H., Li,T., Hu,H., Guan,J.,
            Wu,M., Zhang,R., Zhou,B., Chen,Z., Chen,L., Jin,Z., Wang,R.,
            Yin,H., Cai,Z., Ren,S., Lv,G., Gu,W., Zhu,G., Tu,Y., Jia,J.,
            Zhang,Y., Chen,J., Kang,H., Chen,X., Shao,C., Sun,Y., Hu,Q.,
            Zhang,X., Zhang,W., Wang,L., Ding,C., Sheng,H., Gu,J., Chen,S.,
            Ni,L., Zhu,F., Chen,W., Lan,L., Lai,Y., Cheng,Z., Gu,M., Jiang,J.,
            Li,J., Hong,G., Xue,Y. and Han,B.
  TITLE     Sequence and analysis of rice chromosome 4
  JOURNAL   Nature 420 (6913), 316-320 (2002)
  PUBMED    12447439
REFERENCE   2
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
            Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J., Zhang,L.,
            Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y., Li,C.,
            Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K., Zhou,B.,
            Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P., Fu,G.,
            Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F., Tu,Y.F.,
            Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H., Chen,X.Y.,
            Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,L.J.,
            Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L., Zhu,F.H. and
            Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBb0006N15.
COMMENT     On Jul 9, 2003 this sequence version replaced gi:21912499.
            Web site: http://www.ncgr.ac.cn
            -------------- Summary Statistics
            Assembly program: phrap
            --------------
            This is a complete sequence.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Fgenesh (http://www.softberry.com/),
            genscan (http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM
            (http://genemark.biology.gatech.edu/GeneMark/), tRNAscan-SE (Sean
            Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/), searches of the
            complete sequence against NCBI none redundant protein database (nr)
            (ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.
FEATURES             Location/Qualifiers
     source          1. .133413
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="japonica"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="4"
                     /clone="OSJNBb0006N15"
                     /clone_lib="CUGI-OSJNBb"
     gene            complement(4518. .9189)
                     /gene="OSJNBb0006N15.1"
     CDS             complement(join(4518. .5157,5204. .5367,5563. .6021,
                     6081. .6399,6811. .7089,8338. .8398,9168. .9189))
                     /gene="OSJNBb0006N15.1"
                     /codon_start=1
                     /protein_id="CAE04584.2"
                     /db_xref="GI:38346106"
                     /db_xref="InterPro:IPR005162"
                     /db_xref="UniProt/TrEMBL:Q7XMJ6"
                     /translation="MTSAKPNEGVSLSMRTENESEFAHGKRKEVGVILQPLSSCSTDE
                     IASYLSSPNIGSQAKPEIVNYNNDFDFDNFACYNDYYDDYYGDLNDDYTSLYFGVFMT
                     NNETDEQRAARLADEERTSWGLREQLVHSPAVGNRGPRVDHEQQKSSHGLVDLREALN
                     ARRRARHGRGHHHSTECIDGNDEGVTAFTRDLRRVNWPASFKPTGIEKYDGKTNPESW
                     LIVYTLAIRARNWLAGLLRGTIGSRGELRDHFIANFEGTFERPSTHFELYNVIQKPNE
                     PIRDYIRRFSEKRNKISNMTDGNIITAFTKGVRNELLIGKFGRKPPRTVKEMFEKANE
```

```
                YAKADDAVLASKQSGSGWKPKKDSGNNGASSSNTHHKDRKCKPEDLKKDDTEPAQGTL
                QDSRKELNHIFGGPLAYESKRKQKLTDREINAVQPEVPQFLQCGPPGAVPPVLDPVVR
                NVKLKRTLIDGGRALNILFAKTLDEMQISRIELRLRNAPFHGVFLVFSSTPLGQITMP
                VTFGTGENFHTDNVCFEVADFETAYHAMLGRPALAKFMAIPHYTYMMMKMPRPRGVIS
                LRSDVKQAITCDKESCEMAQSCEQALGREEVCLATSTSPEGDVPAKNLSKSRENDAKT
                KKIPLDPLDPSKTTVIGVELDCK"
gene            complement(16990. .22673)
                /gene="OSJNBb0006N15.2"
CDS             complement(join(16990. .17428,17525. .17570,17651. .19086,
                19291. .19361,19562. .19948,20070. .20225,20915. .21184,
                21764. .21870,22181. .22673))
                /gene="OSJNBb0006N15.2"
                /codon_start=1
                /protein_id="CAE04585.2"
                /db_xref="GI:38346107"
                /db_xref="GOA:Q7X7G6"
                /db_xref="InterPro:IPR000209"
                /db_xref="UniProt/TrEMBL:Q7X7G6"
                /translation="MATPARFSTKPHVFDGTDFSHWCSRMQSYIMAEDYDIWRKVSHA
                YVIPEAINTAAEKTAFEQNCKARNILLSGISRSDYDRVAHLQTAHEIWVALSNFHQGT
                NNIKELRRDLFKKEYIKFEMKPGEALDDYLSRFNKILSDLKYVDSSYDTNYPQSEISR
                HFLNGTGKEVVQIGALNVVLLIIFVRIVLSLGEAKEKTKMLKQPYGVLDCSTCDLNKM
                KLKDALGRVKYMEDVVKNNEVLSCPKCRESQGFMVDCENCANLEKEVSYLKNSFQRFS
                DGKKNLNMILDQSKALVARKENVWIVDSGCSRHMTGDKNWFSSLKKASKTESIIFGDA
                STVPLSLQKTGSKVFDSRGDSVLNISRYGRVFKADFENPVSPVITCLVAKFDKDVMFW
                HRRLGHVGFDHLTRLSGLDLVRGLPKLKKDLELVCTPCRHAKVVSTSHAPIVSVMTDA
                LGQLLHMDTVGPARVHRWRKVVLNMNFLLLVFLNKTVLLKGKFVLNLDKFEARSTDGL
                FLGYPAHTRGYRVLILGTNKIVETCEVSFDEASPGTRPEITGTLSQVQGEDGRIFQDE
                SDYDDDDEVGSAGQTGRQAGQTAGTPPVRPAQDVRSNRPGSSGSGTVDADRDGPPEIT
                TSTSTDTERDSTSEVTAPLHIQRRHPPEQIIGNIGERTTRSKVTAHDVCANSAFVASF
                EPKDVSHALTDESWINAMHEELENFERNKVWTLVEPASGHNIIGIKWVFKNKQNEDGL
                IVRNKARLVAQGFTQVEGLDFDETFAPVARIEAIRLLLAFAASKGFKLYQMDVKSAFL
                NGFIQEEVYVKQPPGFENPDFPNHVFKLSKALYGLKQAPRAWYDRLKNFFLAKGFTMG
                KVDKTLFVLKHGDNQLFVQIYVDDIIFGCSTHALVVDFAETMRREFEMSMMGELSYFL
                GLQIKQTPQGTFVHQTKYTKDLLRRFKMENCKPISTPIGSTAVLDPDEDGETVDQKEY
                RTSPRASYRQAVKRIMRIDRKSTSGTCHFLGTSLIAWSSRKQSSVAQSTTESEYVAAA
                SSCSQIFWLLSTLKDYGLTFEKVPLFCDNTSAINIAKNHVQHSRTKHIDIRFHFLRDH
                VEKGDVELQFLDTKLQIADIFTKPLDSNHFAFLRGELGIIHPFGMV"
gene            30730. .34550
                /gene="OSJNBb0006N15.3"
CDS             join(30730. .33892,34483. .34550)
                /gene="OSJNBb0006N15.3"
                /codon_start=1
                /protein_id="CAE04586.2"
                /db_xref="GI:38346108"
                /db_xref="GOA:Q7FAG3"
                /db_xref="InterPro:IPR000767"
                /db_xref="InterPro:IPR001611"
                /db_xref="InterPro:IPR002182"
                /db_xref="UniProt/TrEMBL:Q7FAG3"
                /translation="MDDIVGILKGAAGLAADANAFREFFNWAAPRVLAAVRSQQQQHL
                LQVDPGGVGSASARDNRTAPLHQIEDDLQNLENDLWLIQSTTSTMYDLIDRLEWHSHK
                DTEDWHLRQIKDVAYDAEDLLDEYNYYALKVKVDTSKNLGQDHSHEPFLEFLNSVNFS
                KVMKIKNRLKQVFDQSKGLGFHKTPKKFDRLVRPETCRVLDEPDEIFGREKELKELKQ
                NLGVRGRKRGRPVACSTTAEARRTELPVLPIVGMGGIGKTTMAQQICEDRVVRNHFDC
                ILWICVSDEFEVNRLTRDVLKSLGVKSQDSDTRDTLMVNLRDSVKSKKFLLVLDDMWD
                DVLKDERGWRTFHRTLSNGLQGSMILVTTRSSKVANLVSNSDPYELNGLQNDVFWDFF
                KLCAFGSNSSRNSPELEHIRPELERIGRAILPKLKGSPLAAKTLGRLLKSNLSVEHWE
                DILRSELWKLEQEETDILPALRLSYVYLPRHMKRCFSICALYPKDHKFDKEFLADIWV
                AQGYVEAEDASSCFNDLVNRSFFQKAAGYSDMYVIHDLIHDTAQLVSKDECFIIQHVS
                DLAKIPSKVRHLSIFTKGNISCTEIVTICTQNKKLRSLICNESYRSIKQFAPVIDCWF
                KELPSIRVLIFKFSTVRKLPESIGNSKHLRYLGLLGSSTFETLPSSVSCLYHVQTINA
                KGCVFKRYPQGFSDLISLNKIESKGFQFNKVKDKQFLEWSTTEIADEQLQMTEEQIEL
                LPHWNLEHLIIKNYLGQSCPSWLQPDCLKVLTSLELYGCRNIQSLSFFDPLFSDLEES
                NSIYHLEVLDVHQCPNIDWQGLVALPSSLRRVTLGNFGHSTDHFVSCFRGLSLLTHLE
                IHCWFLTSIPLQVWMSNLPSLENLRIDNCGSLTSICVSEASNIHTVGVFSSLSTVTIS
                FCNALLSLDEFFMPAYMPFVKTIDVLFCKELALLPVHELHLFSRLEKLCIRNCPKLNM
                QRRMTLPSSLRMLSLLLCPSIEYIDNSHLASAMKLKGLSLQLMRCPDLISIVGAISVS
                EIVQGRIEDCPKLMEITQPFKVMTESQKKVQPDQTMGRLGDENLKN"
gene            52957. .56603
                /gene="OSJNBb0006N15.4"
CDS             join(52957. .55222,55342. .56126,56556. .56603)
                /gene="OSJNBb0006N15.4"
                /codon_start=1
                /protein_id="CAE04587.2"
                /db_xref="GI:38346109"
                /db_xref="GOA:Q7XMJ5"
                /db_xref="InterPro:IPR000767"
                /db_xref="InterPro:IPR001395"
                /db_xref="InterPro:IPR001611"
                /db_xref="InterPro:IPR002182"
                /db_xref="UniProt/TrEMBL:Q7XMJ5"
                /translation="MDDIVGILKGAAGLAADANAFREFFNWAAPRVWDAVRSQQQQHL
                LQVDPRGVGSASARDNRIASLHQIEADLQKLENDLWLIQSTTSTMYDLIDRLKWHSHK
                DTEAWHLRQIKYVVYDAEDLLDEYNYYALKVKVDASKNLGQDHSHETFLEFLDSVNFS
                KVMEIQNRLKQVFDQSKGLGFHKTPKKFDRLVRPETCRVLDKPDKIFGREKELKELKQ
                KLGVCGRKRSRPVACSTIAEARRTELPVLPIVGMGGVGKTTMAQQICEDGVVRKHFDC
                IIWICVSDEFEVNRLTKDVLKSLGVKSQHSDTRDTLMVNLRDSVNAKKFLLVLDDMWD
                DVLKDERGWRIFHRTLSNGLQGSMILVTTRSSKVANLVSNSDPYELNGLQNDVFWDFF
                KLCAFRSNSSSNSPELEHIRPELERIGKAILPKLKGSPLAAKTLGRLLKSNLSIEHWE
                DILRSELWKLEQEETDILPALRLSYVYLPRHMKRCLSICALYPKDHRFEKEFLADIWV
                AQGYVEAEDASSCFNDLVNRSFFQKAAGQNNIYVIHDLIHDTAQLVSKDECFIIQHVS
                DLPKIPSKVRHLSIFTNGNISCTELVTICTQNKKLRSLVCNESYKSIKQFAPVIDCWF
                KELPSIHVLIFKISTVRKLPKSIGNSKHLRYLGLLGSSTFETLPSSVSCLYHVQTINA
                KGCVFKRYPKGFSDLISLKKIESKGFLYNKDKDKQLLEWSTKGITDEQLQMTKEQIEL
                LPHWNLEHLEIVSYLEESKNIYHLEVLKIQNCSNINWQGLVALPSSLRRVTLDKFGHS
                TDHFVSCFRGLTLLTYLEIRCELLTSIPLQVLISNLRSLEDLHFRECSSLTSIGVSEA
                SSRHTVGVFSSLSNVIIMNCDALLSLDEFLMPAYMPVVKTIHVKRCEQLALVPVDELH
                RFSHLEDLCIAYCPKLNMQRRMTLPSSLQMLSLARCPSIKYIDNSHLGSSMTLKGLNL
                ILIRCPDLISILGAISVSEIESGYIDGCPKLMEIKQPFNYKRPRLIIGDVQAMNNDER
                SR"
gene            65512. .67563
                /gene="OSJNBb0006N15.5"
CDS             65512. .67563
                /gene="OSJNBb0006N15.5"
                /codon_start=1
                /protein_id="CAE04588.2"
                /db_xref="GI:38346110"
                /db_xref="GOA:Q7XMJ4"
                /db_xref="InterPro:IPR001878"
                /db_xref="UniProt/TrEMBL:Q7XMJ4"
                /translation="MHSKGIKFMSPLTIDSEERAGEKWPISKFIKNKTLSQPENFISY
                QDSKGNVSIRFTNYETRASDDLDTESSDRNIIDNRRHSISELMEKLDYESEIKYYEDK
                LRDIADEYNNSMKSEWPKIRDKELYFYRELSRIKKQKKEKSLAIETPRTQEIKIEQIK
                LEENQKFDKVLSEEEQWEVNERLLLESYEEENNLIEEIENIEEEDNAEQYNDFIDSLD
                EVGLHNLDNAMEAMEVDLGTGKRKRYGDGFVKNEGERERPARAAGRWPPEKEEYSYNY
                IPGQYRHMGTKRRDFEKPIKFQNQKSDGAILNLAAHDPIDWPNIISIWKGLIVQKYIQ

  Query Match             10.2%;  Score 92;  DB 15;  Length 133413;
  Best Local Similarity   45.2%;  Pred. No. 2.2;
  Matches  405;  Conservative    0;  Mismatches  470;  Indels   21;  Gaps    1;

Qy          3 CGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCG 62
              | | || | |||   |||    |   |   |||    |  || ||  |     ||| |
Db     132201 CCGAGGTCATCCACTTGGGCTCGCCGGTGAAGCGGACGAGGAGCTCCGCCATCGCCGGGG 132260

Qy         63 CGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCG 122
              ||   |   |  | |  | | ||  | |      | ||||  | |   | || ||||||
Db     132261 CGTCGCCGAGGAGGCCGGAGAGCCTCCAGTTCCGGACGAGCACCAGGCGGTCGTCGAGCG 132320

Qy        123 CCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGA 182
              |   | |||| ||  ||| ||  | |   |  |||| ||   |  | ||| |  | |
Db     132321 CGTAGGTGAGGCTGAGGCGGGCGAGGCGCGGGGCGCCGCCGAAGGCCAGCGCGACGGGGC 132380

Qy        183 TGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCA 242
              | |    | ||| | |      ||||| | | || | ||||| ||||||  |  ||
Db     132381 CGTCCACGCACGCCACGCGCTCCAGCGCCGGCGCCGCGCGCAGGTCGACGGTGTGGAAGC 132440

Qy        243 CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGAT 302
              ||    || |  ||||||||||         | | ||     ||||  |  | ||   |
Db     132441 CGCAGCCGTCCACGACGAGCTCGCGGAGCTCCGGCATCCCGTCGACCACCAGCGCGGCGT 132500

Qy        303 CGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCG 362
              |  | ||| |  ||| |    |  |||| |  | |    ||||   ||
Db     132501 ACCACGGCGCGCCGCAGCGGCGGAGGTGGAGCGCCCTGAGCCGCGGGCACGCGTCGAGCA 132560
```

```
Qy        363 GCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCG 422
              || || ||||| |||| |  ||    |  | ||  ||||
Db     132561 CCTGGCGGTAGAAGGCGAGGGGCGTCCGCGGGTGCGTGTCC------------------- 132601

Qy        423 CGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGC 482
              |  | |  ||  | ||  ||  |||| |||| |||                  |  ||
Db     132602 --TGCAGGACGAGCTTGGTGAGCGCGGCGTAGTGCCGGAGCTCGGACGGCGGCGGGAGGA 132659

Qy        483 AGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCA 542
              || | | ||       |||| ||  || |    || || || |  || ||  || |||
Db     132660 AGCAGTTGGACACGGCGAGGCTGCGGATGCGCGACCGGAGCGACTCCGCGCTGAGGCGGT 132719

Qy        543 GCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGA 602
              |    |    |     |||||  | || || |   |||||| |||    | | |   |||
Db     132720 GGTCGGGGAAGCAGAAGGCGTAGCAGGGGCCCTGCTCGACCAGCGCCGGCTTGACGACGG 132779

Qy        603 GCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCT 662
                  ||    | |  |  ||  | | | |   |||| ||   |          |
Db     132780 TGATGTCGAGGTCCTCGACGCCCCACGCGCCGACCGCCATGGCGATCAGGTCCTCGACGA 132839

Qy        663 TGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGG 722
              | | |   |   | |||   | |  |||   |   |  | |   |||  | |  |||
Db     132840 AGCCCGTGCTGCGCGCCGCGAAGAACTCGAGCCGGAGACTCCTGGCGCGGCGCGGGGCGA 132899

Qy        723 TCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGT 782
                |||  |  ||  | ||||     | |  |   | |  ||| |    | | || || |
Db     132900 CGGCGACGCGCGGCTCCATGAAGGCGCTGACGCAGTCGGCGAACCCGCGGATGGCGCGGC 132959

Qy        783 GATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGC 842
              | |||    |  ||   ||      |||||| |  ||| || |||  | ||| |  | ||
Db     132960 GCTCGCACCGGCCGATCGCCGCGTCGAGCTTCCTGGCGTCCGCGCCGGCGGCGTCGCCGT 133019

Qy        843 GGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCG 898
              ||| |||||    |  |  |  |    ||| | | |   || |  ||    || ||
Db     133020 GGGCGCGGACGCAGGCGTGGTAGCGCGGCGGGAGCACGTCGCGGACGCTGAACACG 133075
```

RESULT 4
OSJN00263

```
LOCUS       OSJN00263             148353 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0083D01,
            complete sequence.
ACCESSION   AL731623
VERSION     AL731623.3  GI:32492418
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Feng,Q., Zhang,Y., Hao,P., Wang,S., Fu,G., Huang,Y., Li,Y., Zhu,J.,
            Liu,Y., Hu,X., Jia,P., Zhang,Y., Zhao,Q., Ying,K., Yu,S., Tang,Y.,
            Weng,Q., Zhang,L., Lu,Y., Mu,J., Lu,Y., Zhang,L.S., Yu,Z., Fan,D.,
            Liu,X., Lu,T., Li,C., Wu,Y., Sun,T., Lei,H., Li,T., Hu,H., Guan,J.,
            Wu,M., Zhang,R., Zhou,B., Chen,Z., Chen,L., Jin,Z., Wang,R.,
            Yin,H., Cai,Z., Ren,S., Lv,G., Gu,W., Zhu,G., Tu,Y., Jia,J.,
            Zhang,Y., Chen,J., Kang,H., Chen,X., Shao,C., Sun,Y., Hu,Q.,
            Zhang,X., Zhang,W., Wang,L., Ding,C., Sheng,H., Gu,J., Chen,S.,
            Ni,L., Zhu,F., Chen,W., Lan,L., Lai,Y., Cheng,Z., Gu,M., Jiang,J.,
            Li,J., Hong,G., Xue,Y. and Han,B.
  TITLE     Sequence and analysis of rice chromosome 4
  JOURNAL   Nature 420 (6913), 316-320 (2002)
   PUBMED   12447439
REFERENCE   2
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
            Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J., Zhang,L.,
            Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y., Li,C.,
            Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K., Zhou,B.,
            Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P., Fu,G.,
            Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F., Tu,Y.F.,
            Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H., Chen,X.Y.,
            Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,L.J.,
            Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L., Zhu,F.H. and
            Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2002) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBa0083D01.
COMMENT     On Jul 10, 2003 this sequence version replaced gi:21743050.
            Web site: http://www.ncgr.ac.cn
            -------------- Summary Statistics
            Assembly program: phrap
            --------------
            This is a complete sequence.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Fgenesh (http://www.softberry.com/),
            genscan (http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM
            (http://genemark.biology.gatech.edu/GeneMark/), tRNAscan-SE (Sean
            Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/), searches of the
            complete sequence against NCBI none redundant protein database (nr)
            (ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.
FEATURES             Location/Qualifiers
     source          1. .148353
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="japonica"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="4"
                     /clone="OSJNBa0083D01"
                     /clone_lib="CUGI-OSJNBa"
     gene            complement(5844. .9104)
                     /gene="OSJNBa0083D01.1"
     CDS             complement(join(5844. .5880,5994. .6057,6151. .6224,
                     7164. .7279,7433. .7698,9014. .9104))
                     /gene="OSJNBa0083D01.1"
                     /codon_start=1
                     /protein_id="CAD40684.1"
                     /db_xref="GI:21743051"
                     /db_xref="GOA:Q7X6C1"
                     /db_xref="InterPro:IPR000133"
                     /db_xref="UniProt/TrEMBL:Q7X6C1"
                     /translation="MNAFRLAGDMTHIMSVLVLLLKIHTIKSCAGVSLKTQELYALVF
                     ATRYLDIFTDFISLYNTVMKMIFLGSSFSIVWYIRRHKMVRRSYDKDHDTFRHQFLVL
                     PCFLLALLIHEKFTFREVMWTFSIYLEAVAILPQLVLLQRTRNVDNLTGQYVFFLGAY
                     RALYILNWAYRYFTEPHYVHWITWISGFVQTLLYADFFYYYLNSLKNNVKLTLPD"
     gene            complement(11091. .12734)
                     /gene="OSJNBa0083D01.2"
     CDS             complement(11091. .12734)
                     /gene="OSJNBa0083D01.2"
                     /codon_start=1
                     /protein_id="CAD40685.1"
                     /db_xref="GI:21743052"
                     /db_xref="GOA:Q7XMI3"
                     /db_xref="InterPro:IPR001810"
                     /db_xref="UniProt/TrEMBL:Q7XMI3"
                     /translation="MAAPAGNVETPPRRRVRRRIAAAAADDGGEVDDDGLSVLPDELL
                     LVILGRLDTRTALAAAALSRRLDRLAREIPALVFSVRDVLPPRYHACVRAHGDAAGAD
                     ARKLDAAIGRCERRAIRGFADCVSAFMEPRVAVAPRRARSLRLEFFAARSTGFVEDLI
                     AMAVGAWGVEDLDITVVKPALVEQGPCYAFCFPDHRLSAESLRSRIRSLAVSNCFLPP
                     PSELRHYAALTKLVLQDTHPRTPLAFYRQVLDACPRLRALHLRRCGAPWYAALVVDGM
                     PELRELVVDGCGFHTVDLRAAPALERVACVDGPVALAFGGAPRLARLSLTYALDDRLV
                     LVRNWRLSGLLGDAPAMAELLVRFTGEPKWMTSGPLRSPLPGLRRLVVADMPSNWDVS
                     WPRVILEAAPCLEVLHIHVQEEEDDAGTTAAAAAEIPWPPAESARHERLAELAVVGFA
                     ETRGQVGFVRHVVEACVALRRVVLVRRGRVVVGSEGRYWDWKLVAAEAAAPREGEVGG
                     GGGDGERPTAAAEWSEVEKMVIRSQVKSASRPDVEVFIG"
     gene            complement(15356. .16241)
                     /gene="OSJNBa0083D01.3"
     CDS             complement(join(15356. .15605,16033. .16241))
                     /gene="OSJNBa0083D01.3"
                     /codon_start=1
```

```
                     /protein_id="CAD40686.1"
                     /db_xref="GI:21743053"
                     /db_xref="UniProt/TrEMBL:Q7XVF5"
                     /translation="MVSRSGFGCVSHPRWGMRDRAGKLTEVIEEGDDIGVEEQHGKEG
                     RRRGVWVRILARRWHDFGEVPQMEVWRCGRTTGLAFGGLSWEIKEERSVGGIWLELKK
                     TMRCRRIWRGFGMREGAAGGPRKCLPMRGGGGEGVGNLTDQMYRTRKGYV"
     gene            17718. .20798
                     /gene="OSJNBa0083D01.4"
     CDS             join(17718. .18256,18339. .19127,19307. .19383,20587. .20798)
                     /gene="OSJNBa0083D01.4"
                     /codon_start=1
                     /protein_id="CAD40687.1"
                     /db_xref="GI:21743054"
                     /db_xref="GOA:Q7XVF4"
                     /db_xref="InterPro:IPR001810"
                     /db_xref="UniProt/TrEMBL:Q7XVF4"
                     /translation="MAVRPRRRRRSAPPRAASHSEGDDRLSALPDELLLIVMRGLGTC
                     SALGTDALSRRWARLPRELPALDFRVSDAHPRGYHQRAAEFLALLGGRGDDAAGRKLE
                     ASLARYERRVIRGLAESITGFLDAGAASEERRCVSRLTVEFFATNDAHRIGCVHRLIN
                     TAVGLWGVEELEVAIKPAPLNCPPGVDDAGRAHHSCPGGPAEVDAGFRLRPPLLLLGL
                     PAPEEAPPQEMPLQGPRPGGGRATAIVGGGGGADHGARRRALLVPRHPAEPRRRAGRP
                     GVCIHDANPVVVLFRDVPRLRRVHLSFSLDSSTVDDAQHPLPGPDKYKLDWHVRSERM
                     ASLVVRFTGPERWILPWRVSTRLRSLRRLLVADVPPTWDVSWPRLLLQAAPSLESLHV
                     HVAATADDQPAAAAAPGREIMWPPAMFRHRKLGELVVACFGPTPGQVAFVSQKFLKLP
                     INFGHPGIKSNFEIENPEGTVDDSGNRNCRMHLKSFDMVGSGLIELLHGLPHFLVLLK
                     LSMQRGLLRKTGGMVMLSMSAIFLLLEELE"
     gene            complement(21465. .21785)
                     /gene="OSJNBa0083D01.5"
     CDS             complement(21465. .21785)
                     /gene="OSJNBa0083D01.5"
                     /codon_start=1
                     /protein_id="CAD40688.3"
                     /db_xref="GI:39546272"
                     /db_xref="UniProt/TrEMBL:Q7XVF3"
                     /translation="MADYEVAAHMLGECERRVVEAEICLQAKIDALQMEYDLLAKEVP
                     KKEKEEVRLEKIIVEAGARFKAILVGINMLQRRNEEFLVDSQRYMDEVEPNLKSAITV
                     VEVL"
     gene            23417. .23646
                     /gene="OSJNBa0083D01.6"
     CDS             join(23417. .23495,23591. .23646)
                     /gene="OSJNBa0083D01.6"
                     /codon_start=1
                     /protein_id="CAD40689.3"
                     /db_xref="GI:39546273"
                     /db_xref="UniProt/TrEMBL:Q7XVF2"
                     /translation="MTITELVVERYAPSGHRTESRHRIGKDLIKIRETQGTLSLLLKL
                     "
     gene            29341. .34879
                     /gene="OSJNBa0083D01.7"
     CDS             join(29341. .29472,29559. .29672,29743. .29862,29944. .30570,
                     30644. .30817,30900. .31164,31244. .31314,31398. .31500,
                     31576. .31635,31888. .32099,32185. .32256,32801. .32857,
                     32953. .33237,33714. .33818,34004. .34102,34191. .34322,
                     34393. .34496,34589. .34778,34874. .34879)
                     /gene="OSJNBa0083D01.7"
                     /codon_start=1
                     /protein_id="CAD40690.3"
                     /db_xref="GI:39546274"
                     /db_xref="GOA:Q7XVF1"
                     /db_xref="InterPro:IPR003653"
                     /db_xref="UniProt/TrEMBL:Q7XVF1"
                     /translation="MGHRKRSKVDISSESNSDNDCSLSSDNHDGDVHLSNHDESSDNH
                     DQHFEENDPLLHLTKEELMKKLKIKLGKRPFTHDHKKKRQKVESDSIIDDALLEIHND
                     LVLHSLKMKLSMLGEQSENKRKTFEKDNKCEEVFTRFNVKYFSEVINNLSVHDKEVIG
                     RTCFKSLLNFQSSFVPNQFASWIANHVDVSSSDIVVDDKVIPLTEDCVHIILGLPVGG
                     IEISSNFELGKNKILETFGKSAMPSVKFFGDKFIKGEHMTDDQILISFMLVSLNCFLC
                     PNSSLAPSNKYLSAFENIELIDNLNWSKLIFDWLMKHIRKLEKSKSLGGCFYCLAVNY
                     LDYVNFGLRKLPLDIPRINVWKGNMIKEFSKFDKKSKGVYGRRPLKDISSTCYKMIDT
                     VASSDDVPKKHNNTSFFEMVDSSIPNMLPVDIKNKIHGLLVHYFGNEDDMADERPKKL
                     LVDVLALLADASKSNADTNLSSEVDKLSKSCNPTDQNNDEKNDVTSAHSPKDNMDNDI
                     DDNRNANNKEEIDNVHVAKIMKKLTKEGQEFVTPVNCNPDCNIPSCLAPNKAKSRIVG
                     FNNREPVLSDDDFPKFQIWDPADDIDILNNEVTPVFNHNEKYVVPDSFSPTPAKQAIK
                     KLISTKNLSQTLEDNIEDCVNCSDNEDKENVDTAQQFIKQNSIKKESPDCVIHGERKF
                     SEKCVDLSNQADILYNRLNRVVKIAPEICNESYRSLGRVLLCNMSKSIGHYSQSKSQK
                     DLRRIINPAKYCTDPYTPERQSFQFLHICKFCTITAVKKLLILMEFTANSHLLVLHSN
                     KNDLMKHPSKTNFGMVQKSFSGASLARPIETCDMLFFPILHLRHWFLFVVDLKDETFV
                     FIDSLFEEEEDYQYNARCRLISNFSIVWRKFVPEHPINFASFKIIYPPRPRQTNRIDC
                     GIFTLKFMEIWRPRVMLTNQFSQKDIPNIRIQYVNKLFFHPCNIVLNSTTKKLVTDYY
                     AKG"
     gene            complement(35133. .43646)
                     /gene="OSJNBa0083D01.8"
     CDS             complement(join(35133. .35258,35331. .36318,36614. .37584,
                     37647. .37715,37883. .37960,38058. .38144,38236. .38286,
                     38376. .38447,38532. .38659,39549. .39630,40060. .40109,
                     40398. .40451,40941. .41030,41258. .41338,41798. .41893,
                     42315. .42377,42668. .42757,42985. .43053,43406. .43646))
                     /gene="OSJNBa0083D01.8"
                     /codon_start=1
                     /protein_id="CAD40691.3"
                     /db_xref="GI:39546275"
                     /db_xref="InterPro:IPR004330"
                     /db_xref="InterPro:IPR007527"
                     /db_xref="UniProt/TrEMBL:Q7XVF0"

  Query Match             10.2%;  Score 92;  DB 15;  Length 148353;
  Best Local Similarity   45.2%;  Pred. No. 2.1;
  Matches  405;  Conservative    0;  Mismatches  470;  Indels   21;  Gaps    1;

Qy          3 CGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCG 62
              | | || | |||   |||   |   |   |||   |  || ||  |   |||  |
Db      11633 CCGAGGTCATCCACTTGGGCTCGCCGGTGAAGCGGACGAGGAGCTCCGCCATCGCCGGGG 11692

Qy         63 CGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCG 122
              ||   |  | | | | | ||  | |     | ||||  | |   | || ||||||
Db      11693 CGTCGCCGAGGAGGCCGGAGAGCCTCCAGTTCCGGACGAGCACCAGGCGGTCGTCGAGCG 11752

Qy        123 CCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGA 182
              |  |  |||| ||  ||| ||  | |   |  |||| ||   |  || ||  |  |
Db      11753 CGTAGGTGAGGCTGAGGCGGGCGAGGCGCGGGGCGCCGCCGAAGGCCAGCGCGACGGGGC 11812

Qy        183 TGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCA 242
              | |     | ||| | |      ||||| | | || | ||||| ||||||  |  ||
Db      11813 CGTCCACGCACGCCACGCGCTCCAGCGCCGGCGCCGCGCGCAGGTCGACGGTGTGGAAGC 11872

Qy        243 CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGAT 302
              ||    || |  ||||||||||         |  | ||      ||||  |  | ||  |
Db      11873 CGCAGCCGTCCACGACGAGCTCGCGGAGCTCCGGCATCCCGTCGACCACCAGCGCGGCGT 11932

Qy        303 CGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCG 362
              |  | |||  |  ||| |   | |||| |   |  |  |||   ||
Db      11933 ACCACGGCGCGCCGCAGCGGCGGAGGTGGAGCGCCCTGAGCCGCGGGCACGCGTCGAGCA 11992

Qy        363 GCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCG 422
              || || ||||| ||||  |  ||     |  | ||  ||||
Db      11993 CCTGGCGGTAGAAGGCGAGGGGCGTCCGCGGGTGCGTGTCC------------------- 12033

Qy        423 CGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGC 482
                |  |  ||  | ||  ||  |||| |||| |||                  |  ||
Db      12034 --TGCAGGACGAGCTTGGTGAGCGCGGCGTAGTGCCGGAGCTCGGACGGCGGCGGGAGGA 12091

Qy        483 AGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCA 542
              || | | ||      ||||  ||  || |    || || || |  || ||  || |||
Db      12092 AGCAGTTGGACACGGCGAGGCTGCGGATGCGCGACCGGAGCGACTCCGCGCTGAGGCGGT 12151

Qy        543 GCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGA 602
              |   |   |     |||||  | || || |  ||||| |||    | | |   |||
Db      12152 GGTCGGGGAAGCAGAAGGCGTAGCAGGGGCCCTGCTCGACCAGCGCCGGCTTGACGACGG 12211

Qy        603 GCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCT 662
                ||    | |  |  ||  | | | |   |||| ||   |          |
Db      12212 TGATGTCGAGGTCCTCGACGCCCCACGCGCCGACCGCCATGGCGATCAGGTCCTCGACGA 12271

Qy        663 TGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGG 722
              | |  |   |  | |||  | |  |||  |  |  | |   |||  | |  |||
Db      12272 AGCCCGTGCTGCGCGCCGCGAAGAACTCGAGCCGGAGACTCCTGGCGCGGCGCGGGGCGA 12331
```

```
Qy        723 TCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGT 782
              ||| |  ||  | ||||     | |  |   | |  ||| |     | | || || |
Db      12332 CGGCGACGCGCGGCTCCATGAAGGCGCTGACGCAGTCGGCGAACCCGCGGATGGCGCGGC 12391

Qy        783 GATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGC 842
              | |||    |  ||   ||      |||||| |  ||| || |||  | ||| |  | ||
Db      12392 GCTCGCACCGGCCGATCGCCGCGTCGAGCTTCCTGGCGTCCGCGCCGGCGGCGTCGCCGT 12451

Qy        843 GGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCG 898
              ||| |||||    |  |  |  |    ||| | | |   || |  ||     || ||
Db      12452 GGGCGCGGACGCAGGCGTGGTAGCGCGGCGGGAGCACGTCGCGGACGCTGAACACG 12507


RESULT 5
AF411573S3/c
LOCUS       AF411573S3              1784 bp    DNA     linear   BCT 27-SEP-2001
DEFINITION  Actinomadura verrucosospora polyketide synthase 1 gene, partial
            cds.
ACCESSION   AF411575
VERSION     AF411575.1  GI:15788317
KEYWORDS    .
SEGMENT     3 of 4
SOURCE      Actinomadura verrucosospora
  ORGANISM  Actinomadura verrucosospora
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptosporangineae; Thermomonosporaceae; Actinomadura.
REFERENCE   1  (bases 1 to 1784)
  AUTHORS   Tsantrizos,Y.S. and Yang,X.
  TITLE     Cloning and characterization of a gene cluster involved in the
            biosynthesis of verucopeptin, a cyclodepsipeptide metabolite of
            Actinomadura verrucosospora
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1784)
  AUTHORS   Tsantrizos,Y.S. and Yang,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2001) Department of Chemistry, Boehringer
            Ingelheim (Canada), Laval, Quebec H7S 2G5, Canada
FEATURES             Location/Qualifiers
     source          1. .1784
                     /organism="Actinomadura verrucosospora"
                     /mol_type="genomic DNA"
                     /strain="Q886-2"
                     /db_xref="taxon:46165"
     CDS             <1. .1784
                     /codon_start=3
                     /transl_table=11
                     /product="polyketide synthase 1"
                     /protein_id="AAL07758.1"
                     /db_xref="GI:15788321"
                     /translation="ILIHAAASGVGMAAVQLAQHAGAEVYATASPAKWPVLRALGLDD
                     EHLASSRDLAFADEFLEATGGRGVDVVLNSLAHAFVDASLRLLPEGGHFLEMGKTDIR
                     DPRRVAADHPGVDYRAFDVYGSAGPELLQEMTRAVMDLFADGSLRLNPVSVRDVRDAR
                     KAFREMSHGRHVGKLVLDVGGGFGGGTVLVTGGTGGVGSLVARHLAAEHGVRSLVLAS
                     RRGTAAGGVRELVADLEAAGAAVTVVACDVADRAAVADLIADMPARYPLTAVVHAAGA
                     VADGTVESLTAEDLDHALRAKVGGAINLHELTKGHALSAFILFSALAGALGAGGQANY
                     AAANTFLDGLAARRRASGLACTSLSWGWWERRGGMTAGLDEARLARLRRLGAVEMPAS
                     EALALFDAACDHGGPLLIPARVDLGALRGAARGELPPVLRDLVHGGRPSRGRTGQTGE
                     GPRGLTARLAALADGEAEAALLDWVREQTAIVLGHTSGTSVDADRAFTHLGFDSLTSV
                     ELCNRVASSTGLRLPSTLVFSYPTPRELGRHILGLLRPASAAPGAGPSGGAEAGERAP
                     DPDARAGGLADLDLEGLLDLALDEKGK"
ORIGIN

  Query Match             9.5%;  Score 85.6;  DB 1;  Length 1784;
  Best Local Similarity   44.3%;  Pred. No. 39;
  Matches  395;  Conservative    0;  Mismatches  494;  Indels    3;  Gaps    1;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              | ||| || | | | ||       | | | || ||||| |||  | |  |||| ||  | |
Db       1118 GCCGGCGGTCATGCCGCCCCGCCGCTCCCACCAGCCCCAGCTCAGCGAGGTGCAGGCCAG 1059

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              ||| |  |    | ||||| | |     | | | |   |    || ||||  |
Db       1058 GCCGGACGCCCTGCGCCGCGCCGCCAGGCCGTCCAGGAAGGTGTTGGCCGCGGCGTAGTT 999


Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              |||| ||| |   ||      | || |   | | | |   ||  | | ||  ||
Db        998 CGCCTGGCCGCCCGCTCCGAGCGCGCCCGCGAGGGCGGAGAACAGGATGAACGCCGACAG 939

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              | | | | | | | | || | | | || | | | |||| |
Db        938 AGCGTGGCCCTTGGTCAGCTCGTGCAGGTTGATCGCGCCGCCGACCTTGGCGCGCAGCGC 879

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              | | | | ||| ||  |  || || | |    |    |  |||||   || |
Db        878 GTGGTCGAGGTCCTCGGCGGTCAGCGACTCGACCGTGCCGTCCGCGACCGCGCCCGCCGC 819

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              |    || | |  | ||||   | | | | |   |  || ||    | |||   |
Db        818 GTGCACGACCGCCGTCAGCGGATACCGCGCCGGCATGTCCGCGATCAGGTCGGCGACGGC 759

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              | | | |  || | ||| |  | || ||  ||   |  | |  | ||||||  ||
Db        758 GGCCCGGTCCGCGACGTCGCAGGCCACGACCGTGACGGCCGCGCCCCGCCGCCTCCAGGTC 699

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAG---AAGCG 477
              |||| |      |  |  ||  ||  | |||       | ||  ||  | |||     | ||
Db        698 CGCGACGAGCTCGCGGACGCCGCCGGCCGCCGTCCCCCGGCGGCTCGCGAGCACCAGGCT 639

Qy        478 CCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGAC 537
              |||   |  |  ||   |||       | | ||| |  |  |    || ||  | |
Db        638 CCGCACGCCGTGCTCGGCGGCGAGGTGCCGGGCGACGAGCGAGCCCACTCCCCCGGTGCC 579

Qy        538 GCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCG 597
              || |      | ||  |    | ||| | ||| | ||  | || ||||    || | ||
Db        578 GCCCGTGACCAGCACGGTCCCGCCGCCGAAGCCGCCGCCCACGTCGAGGACCAGCTTGCC 519

Qy        598 CACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAG 657
              ||| |     | |       | |||       | ||  |   ||  ||      || | ||
Db        518 GACGTGGCGGCCGTGGCTCATCTCGCGGAACGCCTTGCGCGCGTCGCGGACGTCCCGGAC 459

Qy        658 CAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGT 717
              |     | | | | |    |  || ||  ||||| || ||       |  |||      |
Db        458 CGAGACGGGGTTGAGGCGCAGGCTCCCGTCGGCGAACAGGTCCATGACCGCGCGGGTCAT 399

Qy        718 CGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGT 777
              | |   | | || |   |  | || ||  | | |   || |   | | |
Db        398 CTCCTGGAGCAGCTCGGGGCCCGCCGATCCGTAGACGTCGAAGGCGCGGTAGTCGACGCC 339

Qy        778 GCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGAT 837
              | ||| ||| |   | | ||  | |  || |    |||| || | | |   || |||
Db        338 GGGGTGGTCGGCCGCCACACGCCGGGGATCGCGGATGTCGGTCTTGCCCATTTCGAGGAA 279

Qy        838 GTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCG 889
              ||  | |    | |   |  || |   ||  || |||| | |||
Db        278 GTGCCCGCCCTCGGGGAGCAGCCGGAGGGAGGCGTCGACGAACGCGTGGGCG 227


RESULT 6
CR352334/c
LOCUS       CR352334              159997 bp    DNA     linear   HTG 09-FEB-2005
DEFINITION  Danio rerio clone DKEY-221H20, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR352334
VERSION     CR352334.6  GI:58865138
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 159997)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
COMMENT     On Feb 10, 2005 this sequence version replaced gi:46559209.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK221H20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 156087 bases at least Q40
            Consensus quality: 156644 bases at least Q30
            Consensus quality: 157332 bases at least Q20
            Insert size: 159297; sum-of-contigs
            Insert size: 160605; 1.3% error; agarose-fp
            Quality coverage: 12.78x in Q20 bases; sum-of-contigs Quality
            coverage: 12.68x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    38359: contig of 38359 bp in length
            *    38360    38459: gap of 100 bp
            *    38460    42262: contig of 3803 bp in length
            *    42263    42362: gap of 100 bp
            *    42363    71362: contig of 29000 bp in length
            *    71363    71462: gap of 100 bp
            *    71463    90991: contig of 19529 bp in length
            *    90992    91091: gap of 100 bp
            *    91092   134901: contig of 43810 bp in length
            *   134902   135001: gap of 100 bp
            *   135002   149352: contig of 14351 bp in length
            *   149353   149452: gap of 100 bp
            *   149453   153318: contig of 3866 bp in length
            *   153319   153418: gap of 100 bp
            *   153419   159997: contig of 6579 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159997
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-221H20"
                     /clone_lib="DanioKey"
     misc_feature    1. .38359
                     /note="assembly_fragment:01793
                     fragment_chain:1"
     misc_feature    38460. .42262
                     /note="assembly_fragment:03784
                     fragment_chain:1"
     misc_feature    42363. .71362
                     /note="assembly_fragment:01035
                     fragment_chain:1"
     misc_feature    71463. .90991
                     /note="assembly_fragment:00574
                     fragment_chain:1"
     misc_feature    91092. .134901
                     /note="assembly_fragment:02665
                     fragment_chain:1"
     misc_feature    135002. .149352
                     /note="assembly_fragment:00176
                     fragment_chain:2"
     misc_feature    149453. .153318
                     /note="assembly_fragment:03848
                     fragment_chain:2"
     misc_feature    153419. .159997
                     /note="assembly_fragment:00001.0"
. ORIGIN
```

```
  Query Match             9.2%;  Score 83.2;  DB 14;  Length 159997;
  Best Local Similarity   45.6%;  Pred. No. 12;
  Matches  407;  Conservative    0;  Mismatches  478;  Indels    7;  Gaps    3;

Qy          2 GCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGC 61
              ||| ||||   | |   | | ||      | ||||| | ||    ||| || |    |
Db      42245 GCGCGGGCGGGCGGCGCGTCGCGGCGACGCCGGCCCGGGGCGGCGCGGGCGGCGGCGCGG 42186

Qy         62 GCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGC 121
              |||| || |  | ||    ||    ||| || |  |  ||  ||||| ||   | | |
Db      42185 CGGGGCGGCGGCCGGCGACGCGGGGCGGGGCCGGCCGGCGGCCGCCGCGCCGCCGGGGCC 42126

Qy        122 GCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAG 181
                ||   | |||   ||  | |  ||  | ||  |||  |  |  || ||
Db      42125 CGGGGGGGCGGCGCGGGGCGCCGGGCCGGGGGGTGGGGGGGCCGGCCATGCCCGCGCCCC 42066

Qy        182 ATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGC 241
              || || | ||   |  | || |||||  | ||  |||  | | | | | |
Db      42065 GGCGCGAGCGGGGCCGGGGGGGGGCGCGCGCGCGCCGGCCGGGGCTGGGCCCCGGGCCCT 42006

Qy        242 ACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGA 301
              | | | | | | || |  ||  || |           | | ||  |  |||| || | ||
Db      42005 GCCGGCGCCCGGCCGCCCCGCGGCGGCGGCAGCGGGCGCGGCGCG--CGACACCGCCCGG 41948

Qy        302 TCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGC 361
              |   | | ||  |||||  | |  | |   |        ||| |  |  ||||||||
Db      41947 CCCGTGCGCCCGGCCCGCGCCCGGGGCCCCCGCCGGCGCCCGCCCTCCGGCGCGGGCGCG 41888

Qy        362 GGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGC 421
              |    |||    |  |  |   ||| | |   | || ||     |  | | |  | ||
Db      41887 CGGCGCCCGCGCGGCCGGGGGCGGCCGGGCCCCCCCCCGCCCGGCCGCCGCGCACCCCCC 41828

Qy        422 GCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGG 481
              |||| | ||| ||||   |   | |||| || |      |||||  |  |   |  |||
Db      41827 CCGCCGGGGGGCGGGCGGCGCGGCGGCCGCGGGGGGGGCCGCGCGCCGGGGGCGCCCGCGG 41768

Qy        482 CAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGC 541
                ||| |     |  || |   | |          |||  |||     | ||||
Db      41767 TCCGGGCCGGGGGCGCGGGCCGCGGGCGGCGGCGCGGCCCGCCGCGCCGGACGGCCGCGG 41708

Qy        542 AGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACG 601
              ||| |  |  ||    | || ||| |||  |   |     |  ||    |   |    |
Db      41707 GGCGGGCGCGGCGCGGCGCCGGGGGGGCGCGCCGCCCGCGCCGGCGGCGGGGGGGCGCGG 41648

Qy        602 AGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGC 661
              | | |    ||  | |  || |  |   |  ||| | |  |  | | | || ||
Db      41647 CCGGCGCCGCGGCGCCGCGCGCCCCCGCGCGCCGGCGCCCCCGGGGCGGGCGCGCGCGGC 41588

Qy        662 TTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCG 721
              |   | ||  ||| || |||   |  | |  |  |||  | |   ||| | |
Db      41587 GGCGCCGGC-CGCCGGGGGCGCGGCCCCGGGCGCGCCCGCGCGCGCCGGCGCGGGGCCGG 41529

Qy        722 GTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCG 781
              |   ||  | | |   ||  ||| | |    ||  |||  || ||||| |   |  ||||
Db      41528 GCGGCGGCGGCGGGCGGCGGGGGGCGGGCGCCCGCGGCGGCGCGCGGCGGCCGGCCGCCG 41469

Qy        782 TGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCG 841
              || || | || | ||   |  | | |  |||  || |  |  |||||     ||
Db      41468 ----CGGCGGGGGCCCCGGGCGGGCGGGGGCGGCGGGCGGCGCCCGGCGCGGGGGGCGCG 41413

Qy        842 CGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCC 893
              ||| || |      | ||||| ||||   | |  || || ||||||  || |
Db      41412 CGGCGGGGCGCGGCCGGGGCGCCGGCCCCCCCCGGGGGCTCGGGCGGCGCGC 41361


RESULT 7
PM7G11B/c
LOCUS       PM7G11B                1094 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm7g11.b, sequence tagged site.
ACCESSION   AL685196
VERSION     AL685196.1  GI:19337723
KEYWORDS    STS.
```

```
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 1094)
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .1094
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm7g11.b"
ORIGIN

  Query Match             9.1%;  Score 82.4;  DB 10;  Length 1094;
  Best Local Similarity   44.4%;  Pred. No. 87;
  Matches  387;  Conservative    0;  Mismatches  477;  Indels    7;  Gaps    3;

Qy         38 GAGCCGATCTGGGTGGAGATCGGCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACG 97
              | || |    || ||| |   ||  | |  ||| | |  |||| ||       | ||
Db        985 GGGCGGGCGCGGTTGGCGGCGGGTTCCGCGCGCGGCGCGCGCGGCGGGGCGCCGCCGCNC 926

Qy         98 CGGAGCGTCCGCGAGGCGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTC-GAGGTCGAG 156
              |   ||  | |||  ||| || |||| ||||   | |||   |||  | | |  |    |
Db        925 CCNCGCCGCNGCGGCGCGGCGCGCGCGCGGCNCGGCGCGGGGCGGGGGGCTGGCGNNCNG 866

Qy        157 GCCGGGGAGCTCGCGCACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTC 216
              |||| | ||  |||  |   |  |  ||    ||||     |  |||| |||||| |
Db        865 NCCGGCGGGCGGGCGNGCCGTGGCGCCGCGNCNCCGCCGCGCGGTGCGGGGCGCGGGCGG 806

Qy        217 CCCCCGCAGATCGACGCGGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTC 276
              |          | || |          ||| |||||   || ||  || |  | ||
Db        805 GGCGGCGCCGCGGGCGGGCGGCGCGGGCGCCGCGCCGGGCGCGCGCGGCGGCGGGGCGCG 746

Qy        277 CCTCAGCCGGACCGACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGC 336
              |  | ||  | |   | |  |        |          | |||| || ||    || | |
Db        745 CGGCGGCGCGCCGCGCNCNCCCGCGGGCCNCGCGGGCGGGCGCGCGCGGCGGCGGGCCNC 686

Qy        337 CGCGTCGCTCCACACCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAG 396
              |||| ||  |  | ||  |||||  |   ||  |   ||| |  |  |||||  || |
Db        685 CGCGGCGGGCGCCGCCCGGGGCGCGGCNCCGGGGCGCCGGGGGGCGGGCGCGGNNGCGCG 626

Qy        397 CTC-----GTCCCATGTCGCCTCGCGGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCG 451
              | |     | | |  | | |  |||||   | || || | |||   |   | | ||| |
Db        625 CCCTCGGGGGCGCGCGGNGGCGGGCGGCCGCCCCCGGCCGGCGGGCGGGCCCCCCGCGGG 566

Qy        452 TGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGG 511
              |  |  ||  ||||  |  |  | | |  |    | |  ||      | ||| |  |
Db        565 GGCGGGCGGGGGCGCGGGNGGCCGGGGCNCCGCCCCCCCCCGGCGCGGCGCGGGGGGCCC 506

Qy        512 GCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGG 571
              |     || |   |||   |     |   |  | |  ||  ||||| ||  ||
Db        505 CCNCNNNGGGGGGGGGCGCGGGGCGGGGCGGGGCGGGGCGCGGGCGGGGCGCNGGCCCGC 446

Qy        572 TCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCCCGGAACAGCGCCTC-TGCCTCCTC 630
              ||     | ||  ||     |  | |   | || | |   | | |  ||  || | |
Db        445 CCGCGGGCCGCGGGCGGCGGGGGGGGGGGGGGCGCGGGCGGCCGGGGGTCGCGCGCCCGC 386

Qy        631 CCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGC 690
              |||| | || |||         |   | | ||    | | |  ||   |
Db        385 GCGGGCGCCCGCGGGGGCCCGGGGGGCGGGGGGGGCGGGGGGGGGNNGGGGGGGGGGGGG 326

Qy        691 GAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTT 750
              |  | |  | || | || ||||  || |  |    ||  | | |   ||   |
```

```
Db        325 GGGGGGCGCCGGGGGCCCGCGCGCGGCCCGGCGGGCGCGGCCGGGGCGCGCNGCCGCCCC 266

Qy        751 CTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAG 810
              | ||          |||| |   |  | | | | ||  | | ||| | |||     || |
Db        265 GGCGCGGCGGCGCCCCGGCCGCGCGCGGGGGGGGGCGGGGGGGGCGGGCGCGGCGCCGGG 206

Qy        811 CTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAG 870
              |    |  | |  | |||||  | | | | | |  |     ||| |  ||| |
Db        205 GGCCCGCGGGGCCCCCCGCNCGGGGNGGGGGGGCGCGGGCGGGGGCGGGGGGCGGGCCGG 146

Qy        871 CGCGCGGAGCGCGGGCGCGGCCCACGCGTCG 901
              ||   || ||||||||||  ||  | | | |
Db        145 CGGCGGGGGCGCGGGCGCCCCCGCCCCCTGG 115
```

```
RESULT 8
BA000030_80/c
WPCOMMENT
  Sequence split into 91 fragments  LOCUS BA000030 Accession BA000030
     Fragment Name      Begin      End
     BA000030_00            1    110000
     BA000030_01       100001    210000
     BA000030_02       200001    310000
     BA000030_03       300001    410000
     BA000030_04       400001    510000
     BA000030_05       500001    610000
     BA000030_06       600001    710000
     BA000030_07       700001    810000
     BA000030_08       800001    910000
     BA000030_09       900001   1010000
     BA000030_10      1000001   1110000
     BA000030_11      1100001   1210000
     BA000030_12      1200001   1310000
     BA000030_13      1300001   1410000
     BA000030_14      1400001   1510000
     BA000030_15      1500001   1610000
     BA000030_16      1600001   1710000
     BA000030_17      1700001   1810000
     BA000030_18      1800001   1910000
     BA000030_19      1900001   2010000
     BA000030_20      2000001   2110000
     BA000030_21      2100001   2210000
     BA000030_22      2200001   2310000
     BA000030_23      2300001   2410000
     BA000030_24      2400001   2510000
     BA000030_25      2500001   2610000
     BA000030_26      2600001   2710000
     BA000030_27      2700001   2810000
     BA000030_28      2800001   2910000
     BA000030_29      2900001   3010000
     BA000030_30      3000001   3110000
     BA000030_31      3100001   3210000
     BA000030_32      3200001   3310000
     BA000030_33      3300001   3410000
     BA000030_34      3400001   3510000
     BA000030_35      3500001   3610000
     BA000030_36      3600001   3710000
     BA000030_37      3700001   3810000
     BA000030_38      3800001   3910000
     BA000030_39      3900001   4010000
     BA000030_40      4000001   4110000
     BA000030_41      4100001   4210000
     BA000030_42      4200001   4310000
     BA000030_43      4300001   4410000
     BA000030_44      4400001   4510000
     BA000030_45      4500001   4610000
     BA000030_46      4600001   4710000
     BA000030_47      4700001   4810000
     BA000030_48      4800001   4910000
     BA000030_49      4900001   5010000
     BA000030_50      5000001   5110000
     BA000030_51      5100001   5210000
     BA000030_52      5200001   5310000
```

```
BA000030_53       5300001   5410000
BA000030_54       5400001   5510000
BA000030_55       5500001   5610000
BA000030_56       5600001   5710000
BA000030_57       5700001   5810000
BA000030_58       5800001   5910000
BA000030_59       5900001   6010000
BA000030_60       6000001   6110000
BA000030_61       6100001   6210000
BA000030_62       6200001   6310000
BA000030_63       6300001   6410000
BA000030_64       6400001   6510000
BA000030_65       6500001   6610000
BA000030_66       6600001   6710000
BA000030_67       6700001   6810000
BA000030_68       6800001   6910000
BA000030_69       6900001   7010000
BA000030_70       7000001   7110000
BA000030_71       7100001   7210000
BA000030_72       7200001   7310000
BA000030_73       7300001   7410000
BA000030_74       7400001   7510000
BA000030_75       7500001   7610000
BA000030_76       7600001   7710000
BA000030_77       7700001   7810000
BA000030_78       7800001   7910000
BA000030_79       7900001   8010000
BA000030_80       8000001   8110000
BA000030_81       8100001   8210000
BA000030_82       8200001   8310000
BA000030_83       8300001   8410000
BA000030_84       8400001   8510000
BA000030_85       8500001   8610000
BA000030_86       8600001   8710000
BA000030_87       8700001   8810000
BA000030_88       8800001   8910000
BA000030_89       8900001   9010000
BA000030_90       9000001   9025608
Continuation (81 of 91) of BA000030 from base 8000001 (BA000030 Streptomyces avermitilis

  Query Match             9.1%;  Score 82.4;  DB 1;  Length 110000;
  Best Local Similarity   45.9%;  Pred. No. 16;
  Matches  410;  Conservative    0;  Mismatches  466;  Indels   18;  Gaps    3;

Qy          8 GCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGC 67
              ||   | ||  ||     |  || ||     | |||| || |  ||  |||| |    |  |
Db      97164 GCGAGCGGGAGGGCCTTGAGCGTCGCCATCAGGCCGGTCAGATTGCCGATCAGGCGGCCC 97105

Qy         68 CGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGG 127
                ||    |||||  |||  | |  | |  ||  | ||  |   |  | || || |  |
Db      97104 GACTTGCCGCGCGCCAGCTCGGCGATGTCCGGGTTCTTCTTCTGCGGCATGATCGACGAG 97045

Qy        128 CAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCG 187
              | |  || ||    |||| |||| || |  || || |||   |  ||   |||||  |
Db      97044 CCGGTGGAGAAGGCGTCGTGGAGGGTGACGAAGGAGAACTCCTTCGTGTTCCAGATGATG 96985

Qy        188 TGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCC 247
                | | ||  ||  |||     | | ||||   ||   ||  |  || |   |   |
Db      96984 ACCTCCTCGGCGATCCGGGAGAGGTTCACCCCAATCATCGCGGTGATGAAGGCGAACTCG 96925

Qy        248 TCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGG 307
               || ||  | |  |    ||| |||||||  |      | |   ||   || | |||  |
Db      96924 GCGACGAAGTCACGGGAGGCCGTGCCGTCGATGGAGTTGGCGACCGAGCCGTGCTCGAAG 96865

Qy        308 ATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCG 367
                | |  | || | |  || | |    | ||     |      |||   | |   |
Db      96864 CCGAGGTCCTTCGCGACCGACTCCGGGTCGAGGCCGAGGGAGGAACCGGCGAGGG---CA 96808

Qy        368 CCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCC 427
              ||  ||  |       || |||||  | ||||||||||       |   ||    ||| ||
Db      96807 CCCGAGCCGTACGGCGACACGGCCGTGCGCTCGTCCCACTGGCGCAGCCGCTCGGCGTCC 96748

Qy        428 GCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAAC 487
```

```
                || | |||| | |    ||||| ||     ||  |      ||  |    || ||  | |
Db      96747 CGGGACAGCGCCTGGACGTGCGCGAGGACGTGGTGGGCGAAGAGCACCGGCTGGGCGTGC 96688

Qy        488 TCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAG 547
              | |  ||     | |   |    | || |  |    |    ||||||  | |   | | |
Db      96687 TGGAGGTGCGTGCGGCCGGGCATCGCCACGTCCGGGTGGGCCTCCGCCAGTCCGATCAGC 96628

Qy        548 AACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCC 607
                | |   | ||  |  | || ||      |||| |   |  | | | |||||  |  |
Db      96627 GCGTCCTGGAGGTCGGCGAGGAGGCCGCCGATGACGCGGGCGTGGTCGCGCAGGTACATG 96568

Qy        608 CGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCG 667
              ||||| || | | |  |||       ||  |  |  ||     |     | | ||||||  ||
Db      96567 CGGAAGAGGGTCGCGACCT------GGTCGTTCCGGGAGCGGCCGGCGCGGAGCTTGCCG 96514

Qy        668 ATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACG 727
                | |  |||           |||  | | ||  ||   | |||||     |||||   |
Db      96513 CCGAGGTCGGG---------GCCGAGGCGCTCCAGCAGGCCGCGCTCCAGGGCGGTGTGG 96463

Qy        728 ATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCG 787
              | ||||   |    |     |  | |   ||   ||  ||    |  ||   || | |  |
Db      96462 ACGTCCTCGTCGGCGATGGTGCCCACGAAGGTGCCGTCCGCGACATCGGCTTCGAGCCGG 96403

Qy        788 ACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGG 847
              ||||  |   |   |    |  | ||||| | || | ||||  |       | | | |
Db      96402 TCGAGGCCCGCGAGCATACGGTCCAGCTCGTCGTCCGTGAGGAGGTTCGCCTTGTGCAGC 96343

Qy        848 CGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTCG 901
                | || || ||||    |     ||| |     |||||||  || |  ||||
Db      96342 ACGCGCGCGTGGGCACGCGAACCTGCGATGTCGTAGGGCGCGAGCCGCCAGTCG 96289


RESULT 9
AR352525/c
LOCUS       AR352525               11604 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 13 from patent US 6589775.
ACCESSION   AR352525
VERSION     AR352525.1  GI:33757733
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11604)
  AUTHORS   Jensen,S.E., Aidoo,K.A. and Paradkar,A.S.
  TITLE     DNA sequence encoding enzymes of clavulanic acid biosynthesis
  JOURNAL   Patent: US 6589775-A 13 08-JUL-2003;
            The Governors of the University of Alberta; Edmonton;
            EPX;
FEATURES             Location/Qualifiers
     source          1. .11604
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             9.0%;  Score 81.2;  DB 6;  Length 11604;
  Best Local Similarity   45.1%;  Pred. No. 47;
  Matches  342;  Conservative    0;  Mismatches  413;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||     |   || || || ||     | ||| |     |  |     |  || | ||
Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |    |  | ||| ||| |    ||||  |       ||| |||  ||
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               ||| ||| | |  |   | ||| |     ||  | |  |     |||| ||  | |
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
                |  ||||      |||  | || || |  | ||| |||  || ||      ||
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842
```

```
Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |     |
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||         || |  ||  |   || ||  |  |  | |||||||  | |||
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |     ||||   | |  | ||| ||   |  |   | ||  | | ||| |  || ||
Db        601 GGAGCTCCACGGCGTACAAGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||      |  |||  ||  | | ||   | |||  |||  |      |
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |        || |  |     |||    |||  |    |   |  | |
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||     | |  | |  ||  |   |||||||    || ||  || |
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCG 868
              ||| ||  ||| || ||   ||  ||  | ||  || |
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCGATGGGG 324


RESULT 10
BX571966_01/c
WPCOMMENT
  Sequence split into 32 fragments  LOCUS BX571966 Accession BX571966
     Fragment Name      Begin     End
     BX571966_00            1    110000
     BX571966_01       100001    210000
     BX571966_02       200001    310000
     BX571966_03       300001    410000
     BX571966_04       400001    510000
     BX571966_05       500001    610000
     BX571966_06       600001    710000
     BX571966_07       700001    810000
     BX571966_08       800001    910000
     BX571966_09       900001   1010000
     BX571966_10      1000001   1110000
     BX571966_11      1100001   1210000
     BX571966_12      1200001   1310000
     BX571966_13      1300001   1410000
     BX571966_14      1400001   1510000
     BX571966_15      1500001   1610000
     BX571966_16      1600001   1710000
     BX571966_17      1700001   1810000
     BX571966_18      1800001   1910000
     BX571966_19      1900001   2010000
     BX571966_20      2000001   2110000
     BX571966_21      2100001   2210000
     BX571966_22      2200001   2310000
     BX571966_23      2300001   2410000
     BX571966_24      2400001   2510000
     BX571966_25      2500001   2610000
     BX571966_26      2600001   2710000
     BX571966_27      2700001   2810000
     BX571966_28      2800001   2910000
     BX571966_29      2900001   3010000
     BX571966_30      3000001   3110000
     BX571966_31      3100001   3173005
Continuation (2 of 32) of BX571966 from base 100001 (BX571966 Burkholderia pseudomallei

  Query Match             9.0%;  Score 80.8;  DB 1;  Length 110000;
  Best Local Similarity   45.4%;  Pred. No. 22;
  Matches  332;  Conservative    0;  Mismatches  397;  Indels    3;  Gaps    1;

Qy        165 GCTCGCGCACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCA 224
              |||||  |||   ||  |  |||| ||    || |   |  |   |||  |||   || |
Db      80240 GCTCGTCCACCCGGATGTAGCCGCGCGCATCGGTTTCGACGCCGGCGTGCTCGAGCCCGA 80181

Qy        225 GATCGACGCGGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCC 284
              |||||  ||  |    |||||  | || || |||  |||     |  ||||  |     |
Db      80180 GATCGTCGGTGTTCGGCACGCGCCCGACGGCGAGCAGCAGGTGCGAGCCGCGCACTTCGC 80121

Qy        285 GGACCGACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGC 344
              |  |    ||| ||| |||| ||   |||      | |||||   ||      | ||
Db      80120 GCGCGCCGGCCGCGCAATCGAGGCCGACGGCGATCCCGCCGCCGTCGCGGCGCACGCTCA 80061

Qy        345 TCCACACCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCC 404
              ||  | |||    |  |  || ||  |   |||  | |  |  | || || |  |
Db      80060 GGCAATCGGCGTCGAGTCGAACGTCGATGCCCTCGCCCTCCAGGATCTCGCGCACCGCGT 80001

Qy        405 ATGTCGCCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGC 464
              | | | ||   |  | ||| |||  | ||     || |  | ||  |  |  | ||
Db      80000 GGGACACGTCCTCGTCCTCGCGCGCGATCAGCCGCGGGCCCTTCTCGACGATCGTCACGC 79941

Qy        465 GCCCGAAGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCC 524
              || ||  |||||||||| | | || | || | ||||||   | |  |  |     | |
Db      79940 GCGCGCCGAAGCGCCGGTACATCTGGCCGAACTCGAGCCCCACGTAGCTGCCGCCGAGAA 79881

Qy        525 AGGCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGA 584
              |  |  ||   || |||| |   |                ||| |  || |     |
Db      79880 CGATCAGGTGCTCGGGCAGGAAATCGACATCCATCATCGTCGAGTTCGTGAAATACGGCA 79821

Qy        585 GCCAGAGATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGG 644
              |          || |  || |  | | | ||||||     |  |  |  |  | ||  |
Db      79820 CCTGATCGAGGCCCGGGATCGGCGGAATCAGCGCGCGCCCGCCGACATCGATGAAGATCC 79761

Qy        645 TCTTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCG 704
              || |      |||||||   |||    ||| | || ||  |   |||           |||
Db      79760 GCTCGGCTTCGAGCAGCGCTTCGT---CGACGCGCACCGTGCGCGCGCTCTCGAAGCGCG 79704

Qy        705 TCCTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGA 764
                |  |||   | |   || | ||          || | |||  | |  |
Db      79703 CATGCCCGTGGTAGACCGCGCCGTTCTCGAGACCGCGAACCCATTGCTCGACGCCGCGGT 79644

Qy        765 GCGGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGG 824
              ||    |   ||   | |||||   | ||||     |   |  |  |  | ||| |||
Db      79643 TCGAGCGCGCGGCCACCTGATCCTTGCGCGCCTTCACCTTCTTCATGTCGACCGCCACGG 79584

Qy        825 CGAGGCCGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGG 884
              |  |||| |  |  ||| |     |    ||||    |||  | |||||  |  ||||
Db      79583 GGCCGCCGAGCGTCACGCCGTATTCGTCGGCGCGCCGCGCGAGGTGCGCGGCGTACGCGC 79524

Qy        885 GCGCGGCCCACG 896
              ||||  | | |
Db      79523 TCGCGATCAAGG 79512


RESULT 11
AF533768
LOCUS       AF533768               156789 bp    DNA     linear   VRL 19-MAY-2003
DEFINITION  Cercopithecine herpesvirus 1 strain E2490, complete genome.
ACCESSION   AF533768
VERSION     AF533768.1  GI:30844242
KEYWORDS    .
SOURCE      Cercopithecine herpesvirus 1 (CeHV-1)
  ORGANISM  Cercopithecine herpesvirus 1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Simplexvirus.
REFERENCE   1  (bases 1 to 156789)
```

```
  AUTHORS   Perelygina,L., Zhu,L., Zurkuhlen,H., Mills,R., Borodovsky,M. and
            Hilliard,J.K.
  TITLE     Complete sequence and comparative analysis of the genome of herpes
            B virus (cercopithecine herpesvirus 1) from a rhesus monkey
  JOURNAL   J. Virol. 77 (11), 6167-6177 (2003)
   PUBMED   12743273
REFERENCE   2  (bases 1 to 156789)
  AUTHORS   Perelygina,L., Zhu,L., Zurkuhlen,H. and Hilliard,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) Department of Biology, Georgia State
            University, 24 Peachtree Center Ave, Atlanta, GA 30303, USA
FEATURES             Location/Qualifiers
     source          1. .156789
                     /organism="Cercopithecine herpesvirus 1"
                     /mol_type="genomic DNA"
                     /strain="E2490"
                     /specific_host="Macaca mulatta"
                     /db_xref="taxon:10325"
     misc_feature    1. .223
                     /note="a sequence"
     repeat_region   224. .9021
                     /note="terminal copy of large repeat region"
     gene            2194. .4636
                     /gene="RL2"
     CDS             join(2194. .2241,2461. .3003,3125. .4636)
                     /gene="RL2"
                     /note="modulator of cell state and gene expression"
                     /codon_start=1
                     /product="immediate early protein ICP0"
                     /protein_id="AAP41493.1"
                     /db_xref="GI:30844317"
                     /translation="MEPSRGLGAPARPAAEAPAEDALEDDDLVLLSRYFFPDTDSDDD
                     DNDDDAAAVVDDDPASDAAAADGTGLSAPAPPRETCAVCTERIDEAQLCNTFPCLHRF
                     CIPCLKTWLPMRNSCPLCNAAVSYLIVGVQANGSYSTIPVVNDPRTRRDAEDAVRAGT
                     AVDFIWVDRPGEPAPSSVTLGGRTVRALSPPARMGRPGVVGRVSRPPRSPPPPPAPPA
                     PDSPPASLSLRPSADPTFHPAPARGARAPAARGDAARAHRASTRGGARGGARGAAAPR
                     APAPAGNPAPIVIGDSPPASPRRPSGAAAPPVVPVAPRPRPAAEPPAPAPGAAPRQPA
                     HQTQAQARAQARTQARAQAVLAQALAQALTQARPGAQAQPRPGAQVPPGAQVPPGAQV
                     PPGAQVPPGAQAQPRPAAQAQPAAQVPPGAQARPQARARPAAPPQAPPLCAVVAAAAA
                     ARKRPASASEAAPATRGPKRPSLPPPPGGPAPPRRAPARAPALAAAPPPPRPLAPPAA
                     PPPPQAAPPPPAAPPPPPVETRDRSSLGPRPAEQGPRKCVRKTRHVDAEGAPAAPGPT
                     RYLPISGVSSVVAMAPYLNKTVTGDCLPVLDMETGDIGAYVVLVGRACNMARALAEAS
                     PGWSRRTRLPEAAPGRVSPPEYPGDPAHGLWMTPVGGMLFDQGALLGGRSFHSLDSRH
                     PWTPGPADPPPTRGSGGP"
     exon            2194. .2241
                     /gene="RL2"
                     /number=1
     exon            2461. .3003
                     /gene="RL2"
                     /number=2
     exon            3125. .4636
                     /gene="RL2"
                     /number=3
     misc_feature    9022. .116838
                     /note="unique long region"
     gene            9072. .9746
                     /gene="UL1"
     CDS             9072. .9746
                     /gene="UL1"
                     /note="proposed start codon is GTG"
                     /codon_start=1
                     /transl_except=(pos:9072. .9074,aa:Met)
                     /product="virion glycoprotein L"
                     /protein_id="AAP41419.1"
                     /db_xref="GI:30844243"
                     /translation="MFLRSVSAAPSVVSPAASPSPSPPVEYVIRSVAARTVGDILKFA
                     CLELPAGGVTWRYEAPRSIDYARIDGIFLRYHCPGLDTVVWDGKAQRAYWVNPFLFAA
                     GFLEDLGHALFPANALETTTRFALYKEVRLALASRSDAASSTPVPPGCVDAEYSRTRD
                     CPDGRTPGIWNEPRIRRPFSAPNDEASPQPQSLAPAPTPTPPGRTHEPARKPRGNATR
                     TARPRA"
     gene            9814. .10575
                     /gene="UL2"
     CDS             9814. .10575
                     /gene="UL2"
                     /note="proposed start codon is GTG"
                     /codon_start=1
                     /transl_except=(pos:9814. .9816,aa:Met)
                     /product="uracil-DNA glycosylase"
                     /protein_id="AAP41420.1"
                     /db_xref="GI:30844244"
                     /translation="MTFDSQPPSRPSLGLDDAHADRPPALDWTGFRRRFLVGEAWRPV
                     LEPELSNPLTAQLMAEYERRCRVEEVLPPREDVFSWTRFCTPDEVRVVIIGQDPYHQP
                     GQAHGLAFSVRPGTPTPPSLRNILAAVRNCCPDAELAAHGCLEKWARGGVLLLNTTLT
                     VRRGEPASHAKMGWDRFVASAVRRLAASRPGLVFMLWGAHAQNAIRPDPRVHRVLTYS
                     HPSPLSKVPFSSCRHFCLANQYLRERSLAPVDWST"
gene                 10711. .11394
                     /gene="UL3"
CDS                  10711. .11394
                     /gene="UL3"
                     /codon_start=1
                     /product="UL3"
                     /protein_id="AAP41421.1"
                     /db_xref="GI:30844245"
                     /translation="MSSLEAAAVPSTLAVLASWGWAFAPQDSPRAWAADPAAHTEPEA
                     LALLPAPPRATTDEPPDPRALDADHVAFDTMFMVSSVDELGRRQLTDTIRKDLRLSLV
                     NFAIACTKTSSFSGAAGGARPRDQREPGRREPPSHKSLQMFVLCQRANATRVREQLRA
                     VIASRKPRKYYTRSSDGRISPAVPVFVHEFVSDAPVYIHRDNVIGPPGPPHRTPPRCP
                     EPTDRNTCQ"
gene                 complement(11509. .12120)
                     /gene="UL4"
CDS                  complement(11509. .12120)
                     /gene="UL4"
                     /codon_start=1
                     /product="UL4"
                     /protein_id="AAP41422.1"
                     /db_xref="GI:30844246"
                     /translation="MAASPATIAYSLHGTATSSACVLPDAEQVVCAFDGGTRAIASRG
                     CLRQDTLSRGAVVVRQTPVGLLVMVDCRTDFCAYRFVGRSGRQRLERWWDTNLCAYPF
                     DSWVSSTRGESVRSPTAGIVTVVWGEDSIYITVTVYGSPPGGEAEAPLPDPSSPPLDA
                     PGTPSPPAVSPASAEAGAAADLLVEVMKEIQLSPTLGFGQHEP"
gene                 complement(12172. .14820)
                     /gene="UL5"
CDS                  complement(12172. .14820)
                     /gene="UL5"
                     /codon_start=1
                     /product="component of helicase-primase complex"
                     /protein_id="AAP41423.1"
                     /db_xref="GI:30844247"
                     /translation="MARVGGDASVSSPPEQPREPSVQFRDEAFLNFTSMHGIQPIIER
                     VRELAEHRLEAARVPRLQWFRDVAALETPAGLPLREFPFAVYLITGNAGSGKSTCVQT
                     LNEVLDCVVTGATRIAAQNMYVKLSGAFLSRPINTIFHEFGFRGNHVQAQVGQHPYTL
                     ASNPASVEDLQRRDLTYYWEVIADITRRAVAAGGGAEDGRNEFRALATLESTMGLGEG
                     ALTRLAAFTHGALPAFTRSNIVVIDEAGLLGRHLLTAVVYCWWMTNALYHTPQYAAGL
                     RPVLVCVGSPTQTASLESTFEHQKLRCSIRQSENVLTYLICNRTLREYARLPRSWAIF
                     INNKRCVEHEFGNLMKALEYGLPLTEEHMQFVDRFVVPESFITNPANLPGWTRLFSSH
                     KEVSAYMARLHAHLKVAREGEFVVFSLPVLTFVSAKAFDEYRELTHQPALTVEKWIAA
                     NAGRITNYSQSQDQDAGRARLEVHSKQQLVVARNDVTYVLNSQIAVTTRLRKLVFGFS
                     GTFRSFEAVLRDDSFVKTQGEMLVEFAYQFLSRLIFGGLIHFYNFLLRPGLDAARRAR
                     AYERMGALTRELLSLGATGAGPATPGDQAASGGGESASAAPGAERAFDFKTLGARATD
                     AEFPEDDLDVVFAGLNEQELDLFYCHYTVEQPETTAAVHAQFGLLKRAFLGRYHILRE
                     LFGEAFEGAPFGTYVDNVTFRGCEVMIGSMRGGLMSVALQTDSYTLTGFTYARVYAFA
                     EEVRRRHAAGGLAELLEESPLPYVVLRDQHGFMSVVNTNISEFVESVESRELAMATNA
                     DYGISSMLAMTITRSQGLSLDKVAICFTPGNLRLNSAYVAMSRTTSSEFLRMNLNPLR
                     ERHERDDVISEHILSALRDPDVLIVY"
gene                 14819. .16870
                     /gene="UL6"
CDS                  14819. .16870
                     /gene="UL6"
                     /note="role in DNA cleavage/packaging"
                     /codon_start=1
                     /product="capsid protein"
                     /protein_id="AAP41424.1"
                     /db_xref="GI:30844248"
                     /translation="MAERHPRAARPPARPFDAAPRPPDDGWVQVHPTPTTMLFREALL
                     GELGYTEGQGVYNVVRSSEAVTRQLQAAVFGALLNAAAHRDLEADWRRHVAARRLEPE
                     RLVRRYGGAPDGDVAAVAERVYDTWRAALRASLLDFARGLADCFAQGGPGGPVGFSRY
```

```
                     IDWLMCLGLVPVMRRPREGEVTQRLEAFLSDHALPRDLATVAGAVERAAPGLRELARA
                     FDSARIADYDRVCLFYHPRRGEWRVRDPAGAQRGECVVLWPPLWTGDRLLFDSPMQRL
                     SREIVACRALRFHARVCRLRNAASVKVLLGRKNDGDRGAAGAARAVSRALGEDDASKA
                     GSAASRLVRLIINMKGMRHIGDINDTVRAYLDEAGGHLIDSAAVDPALPGFGRATGRP
                     AAGGADPAAPRQQQLRQAFQTSVVNNINGMLEGYINNLFGTIERLRETNTDLATQLRD
                     RDIELRRARLGALDRQQRAADPGRAPEAEPAGQDDLRTDYDIIDVSKAMDDDTYVANS
                     FQHQYVPSYAQDLERLSRLWEHELLRCFKISRHTNNQGRETAIFYSSGAIALFVAPYF
                     ASVLRAPAMGALITGPNGVLGEEELWEAIFKKTRVQTYLTDLAALFVADVRHAARSRP
                     PAPEPADPDRRVRERPRGRSRSPVDAGRPGGRGPEDAWGGRGDHGHLDPRRGGDGGRR
                     P"
     gene            16821. .17711
                     /gene="UL7"
     CDS             16821. .17711
                     /gene="UL7"
                     /note="role in DNA cleavage/packaging"
                     /codon_start=1
                     /product="capsid protein"
                     /protein_id="AAP41425.1"
                     /db_xref="GI:30844249"
                     /translation="MATSTPAEEETAAAALRRAVAGDRDLVGVAEAISHQALLRMACE
                     VRQATRHTVRFTATGVVRVNVTPDHRLHFVLDGCPDDALIASDAYFARCRDQSAHRGF
                     AFAVVLAREDHVHSLCVPPLVLPHRLSVFWPEHLEDFELACLLMYLENCPRAHATASA
                     FVKISAWLAALGRRTSPFDRVRCLLLRSCQWMLNTLMFMVHVEPFDDQYVLPHWCMAR
                     YLLANNPPPILTALFRAGPSRVFRLPVQPRASLECVAYNPQGVLAAGWASEDFRAGLV
                     SWWSSDAPKTQTSALFYRLV"
     gene            complement(17886. .20171)
                     /gene="UL8"
     CDS             complement(17886. .20171)
                     /gene="UL8"
                     /codon_start=1
                     /product="component of helicase-primase complex"
                     /protein_id="AAP41426.1"
                     /db_xref="GI:30844250"
                     /translation="MGDAGLVWAEESICAITLYTAWLPPRTRDRLHVLLYLFCRDARG
                     DGAARFAEVSIASTELQAFYGPAEISAAGAVAAARAAASPAAAPLETLNDPVLWRALY
                     ACALVALEREVGRAALFAPARVAWDPRTGLVARIEGLPDGRSPAPRAAALDVNAQIAT
                     DPLALAAHVAARPDARLAWARLAAIQDTPQCASAASLTVRIDTGAARFAREYTTLTFP
                     PVRREGAVADVFEVREAVLRPRGHAQAVTARVLVPRGYEYFAVGAEGFSAPALIAMFR
                     QWYLTVHAGPGALPPIFAFLGPEFEPRGGAVDHFSVLGFPGWATLKVSGGAAAAPPAA
                     LAAHAELAGAWPAAGARALGPARAWAAVSAAAAERLPPALRRPAAAGLRLLDPPAVIG
                     PVCLARFRFPGLQPRLLAALYELGLRGDPRAADPRDPLGRGARRWRDLAGGEGAAEQA
                     PWPREVLALATDEVRREYPAVARTLDGIVSAVGVGLEKAAGAVDMAVCGDGGGRVWGV
                     LNVDPRDAAGAGAAAAAAERARAALAASASEALAGWGLRLDAPPPLVLEGTYTHAVLW
                     SQTGAWFWNSDTDEDRLEGFPLRGPAYAAAAAVARRALRAVVAAGSEGGAGAEEALRI

  Query Match             8.9%;  Score 80.2;  DB 13;  Length 156789;
  Best Local Similarity   44.6%;  Pred. No. 22;
  Matches  362;  Conservative    0;  Mismatches  443;  Indels    6;  Gaps    1;

Qy         52 GGAGATCGGCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGA 111
              || |  | |||||||| || | ||      |          | ||| ||  | |   | |||
Db     127973 GGCGCGCAGCGCGGGCGGCAGGGCGTCGGGCCCGGCGCGCGCGGCGTCGCGGTCCAGCGC 128032

Qy        112 GGCGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCG 171
                |  ||  | |||     | |||  | |  |   | |  ||  | || || ||  | |
Db     128033 CAGGTAGACGGGCCGCAGCGGCGCGGCCAGCCCCCAGCGCGCGCAGGCGCGGTGCGAGTG 128092

Qy        172 CACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGAC 231
              | | | |  | || |     || |  | |  |  |||||||| |       |   | |
Db     128093 CGCCTCGCCGCGGCGGAAGTCCGGGGCGCCGGCGGCGGCGGCCTCGCGGGCGGGCGAGTC 128152

Qy        232 GCGGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGA 291
              |  | || |  || |     ||  | |  |       |  || | | || | |  |
Db     128153 GGCGGCGCGGCCGTCGAGCGCGGGCAGCACGGCGCGGCGGTACTCGCGCGGCGGCAGGGG 128212

Qy        292 CGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACAC 351
              | |      ||| |    | ||||    |  |||    |  | |   | |||  |   |
Db     128213 CACGGGCGTGTCGGGCCCGAGGCGCGTGCGCGCGCGGTAGCGCACGTTGGCGCCGCGGCC 128272

Qy        352 CGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGC 411
              |  | ||||||||    |||   ||  |  |  ||||    ||  |||| | |  |
Db     128273 CAGGCGCGGCGGCGGCGCGTCCGGGTAGAGGCGCGCGCAGGCGGCCTCGACGCGCGCGAA 128332

Qy        412 CTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAA 471
              |||  | || ||     ||      | |     |   ||| || |||||  |||| ||
Db     128333 GACGCCGGGCCCGAAGACGCGGCCGGAGGCCAGCACGGCGCGGCGCAGGTCGCGCGCGGC 128392

Qy        472 GAAGCGCCG------GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCA 525
              |    | ||      |||   | |||||      |  | | || |   | |   ||| |
Db     128393 GGGCCAGCGGACGTCGCACTGCGCGGCGGGCAGCACGGCGCAGGGCAGGTAGACGTGGCG 128452

Qy        526 GGCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAG 585
              |  | || || || ||      | | |   |  || |||     | ||  | ||||||
Db     128453 GCGCGCGACGGCGGGCCCGTCGGCCGGCCACTCCTCGGGGGGCACGGCGTCCACGACGAC 128512

Qy        586 CCAGAGATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGT 645
                | |  ||  ||   |  || ||| |   |||  |   |  ||| | ||
Db     128513 GAGGCGGCGGCCGGCGGCGGCGGCGAGCAGGCCGAGGTACTCGACGGCGCCGGCGAAGGC 128572

Qy        646 CTTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGT 705
              |  || |                || ||||    ||||  |      ||  |||  || ||||
Db     128573 CAGGTCCCGCGTGGAGAGCAGGAGCACGCCCTGCGCGCCCAGCGCCGAGACGTCGGGCGG 128632

Qy        706 CCTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAG 765
              | |  |  | |  ||| |  |  | |||| |        |||      | | |
Db     128633 GCCGGTCCAGTTGCCGGCCCAGGCGGCCGTGTCGGGCCCGCAGAGGCGGTTGCCCAGCGC 128692

Qy        766 CGGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGC 825
              |  | | |  |  | | || || ||||| ||     ||| || |  |||     |||
Db     128693 GGCGAGCAGGAAGGAGAGGCCGCCGCGCGCGTCGGTCCAGGCGGGCGGGGCGGGGGAGGC 128752

Qy        826 GAGGCCGGGGATGTCGCGGGGGCGGAGCTCG 856
              |  |||| |    |||| |||  |  || ||
Db     128753 GCCGCCGAGCTCCTCGCCGGGCAGCGGCGCG 128783


RESULT 12
AF533768/c
LOCUS       AF533768               156789 bp    DNA     linear   VRL 19-MAY-2003
DEFINITION  Cercopithecine herpesvirus 1 strain E2490, complete genome.
ACCESSION   AF533768
VERSION     AF533768.1  GI:30844242
KEYWORDS    .
SOURCE      Cercopithecine herpesvirus 1 (CeHV-1)
  ORGANISM  Cercopithecine herpesvirus 1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Simplexvirus.
REFERENCE   1  (bases 1 to 156789)
  AUTHORS   Perelygina,L., Zhu,L., Zurkuhlen,H., Mills,R., Borodovsky,M. and
            Hilliard,J.K.
  TITLE     Complete sequence and comparative analysis of the genome of herpes
            B virus (cercopithecine herpesvirus 1) from a rhesus monkey
  JOURNAL   J. Virol. 77 (11), 6167-6177 (2003)
   PUBMED   12743273
REFERENCE   2  (bases 1 to 156789)
  AUTHORS   Perelygina,L., Zhu,L., Zurkuhlen,H. and Hilliard,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) Department of Biology, Georgia State
            University, 24 Peachtree Center Ave, Atlanta, GA 30303, USA
FEATURES             Location/Qualifiers
     source          1. .156789
                     /organism="Cercopithecine herpesvirus 1"
                     /mol_type="genomic DNA"
                     /strain="E2490"
                     /specific_host="Macaca mulatta"
                     /db_xref="taxon:10325"
     misc_feature    1. .223
                     /note="a sequence"
     repeat_region   224. .9021
                     /note="terminal copy of large repeat region"
     gene            2194. .4636
                     /gene="RL2"
     CDS             join(2194. .2241,2461. .3003,3125. .4636)
                     /gene="RL2"
                     /note="modulator of cell state and gene expression"
                     /codon_start=1
```

```
                     /product="immediate early protein ICP0"
                     /protein_id="AAP41493.1"
                     /db_xref="GI:30844317"
                     /translation="MEPSRGLGAPARPAAEAPAEDALEDDDLVLLSRYFFPDTDSDDD
                     DNDDDAAAVVDDDPASDAAAADGTGLSAPAPPRETCAVCTERIDEAQLCNTFPCLHRF
                     CIPCLKTWLPMRNSCPLCNAAVSYLIVGVQANGSYSTIPVVNDPRTRRDAEDAVRAGT
                     AVDFIWVDRPGEPAPSSVTLGGRTVRALSPPARMGRPGVVGRVSRPPRSPPPPPAPPA
                     PDSPPASLSLRPSADPTFHPAPARGARAPAARGDAARAHRASTRGGARGGARGAAAPR
                     APAPAGNPAPIVIGDSPPASPRRPSGAAAPPVVPVAPRPRPAAEPPAPAPGAAPRQPA
                     HQTQAQARAQARTQARAQAVLAQALAQALTQARPGAQAQPRPGAQVPPGAQVPPGAQV
                     PPGAQVPPGAQAQPRPAAQAQPAAQVPPGAQARPQARARPAAPPQAPPLCAVVAAAAA
                     ARKRPASASEAAPATRGPKRPSLPPPPGGPAPPRRAPARAPALAAAPPPPRPLAPPAA
                     PPPPQAAPPPPAAPPPPPVETRDRSSLGPRPAEQGPRKCVRKTRHVDAEGAPAAPGPT
                     RYLPISGVSSVVAMAPYLNKTVTGDCLPVLDMETGDIGAYVVLVGRACNMARALAEAS
                     PGWSRRTRLPEAAPGRVSPPEYPGDPAHGLWMTPVGGMLFDQGALLGGRSFHSLDSRH
                     PWTPGPADPPPTRGSGGP"
exon                 2194. .2241
                     /gene="RL2"
                     /number=1
exon                 2461. .3003
                     /gene="RL2"
                     /number=2
exon                 3125. .4636
                     /gene="RL2"
                     /number=3
misc_feature         9022. .116838
                     /note="unique long region"
gene                 9072. .9746
                     /gene="UL1"
CDS                  9072. .9746
                     /gene="UL1"
                     /note="proposed start codon is GTG"
                     /codon_start=1
                     /transl_except=(pos:9072. .9074,aa:Met)
                     /product="virion glycoprotein L"
                     /protein_id="AAP41419.1"
                     /db_xref="GI:30844243"
                     /translation="MFLRSVSAAPSVVSPAASPSPSPPVEYVIRSVAARTVGDILKFA
                     CLELPAGGVTWRYEAPRSIDYARIDGIFLRYHCPGLDTVVWDGKAQRAYWVNPFLFAA
                     GFLEDLGHALFPANALETTTRFALYKEVRLALASRSDAASSTPVPPGCVDAEYSRTRD
                     CPDGRTPGIWNEPRIRRPFSAPNDEASPQPQSLAPAPTPTPPGRTHEPARKPRGNATR
                     TARPRA"
gene                 9814. .10575
                     /gene="UL2"
CDS                  9814. .10575
                     /gene="UL2"
                     /note="proposed start codon is GTG"
                     /codon_start=1
                     /transl_except=(pos:9814. .9816,aa:Met)
                     /product="uracil-DNA glycosylase"
                     /protein_id="AAP41420.1"
                     /db_xref="GI:30844244"
                     /translation="MTFDSQPPSRPSLGLDDAHADRPPALDWTGFRRRFLVGEAWRPV
                     LEPELSNPLTAQLMAEYERRCRVEEVLPPREDVFSWTRFCTPDEVRVVIIGQDPYHQP
                     GQAHGLAFSVRPGTPTPPSLRNILAAVRNCCPDAELAAHGCLEKWARGGVLLLNTTLT
                     VRRGEPASHAKMGWDRFVASAVRRLAASRPGLVFMLWGAHAQNAIRPDPRVHRVLTYS
                     HPSPLSKVPFSSCRHFCLANQYLRERSLAPVDWST"
gene                 10711. .11394
                     /gene="UL3"
CDS                  10711. .11394
                     /gene="UL3"
                     /codon_start=1
                     /product="UL3"
                     /protein_id="AAP41421.1"
                     /db_xref="GI:30844245"
                     /translation="MSSLEAAAVPSTLAVLASWGWAFAPQDSPRAWAADPAAHTEPEA
                     LALLPAPPRATTDEPPDPRALDADHVAFDTMFMVSSVDELGRRQLTDTIRKDLRLSLV
                     NFAIACTKTSSFSGAAGGARPRDQREPGRREPPSHKSLQMFVLCQRANATRVREQLRA
                     VIASRKPRKYYTRSSDGRISPAVPVFVHEFVSDAPVYIHRDNVIGPPGPPHRTPPRCP
                     EPTDRNTCQ"
gene                 complement(11509. .12120)
                     /gene="UL4"
CDS                  complement(11509. .12120)
                     /gene="UL4"
                     /codon_start=1
                     /product="UL4"
                     /protein_id="AAP41422.1"
                     /db_xref="GI:30844246"
                     /translation="MAASPATIAYSLHGTATSSACVLPDAEQVVCAFDGGTRAIASRG
                     CLRQDTLSRGAVVVRQTPVGLLVMVDCRTDFCAYRFVGRSGRQRLERWWDTNLCAYPF
                     DSWVSSTRGESVRSPTAGIVTVVWGEDSIYITVTVYGSPPGGEAEAPLPDPSSPPLDA
                     PGTPSPPAVSPASAEAGAAADLLVEVMKEIQLSPTLGFGQHEP"
gene                 complement(12172. .14820)
                     /gene="UL5"
CDS                  complement(12172. .14820)
                     /gene="UL5"
                     /codon_start=1
                     /product="component of helicase-primase complex"
                     /protein_id="AAP41423.1"
                     /db_xref="GI:30844247"
                     /translation="MARVGGDASVSSPPEQPREPSVQFRDEAFLNFTSMHGIQPIIER
                     VRELAEHRLEAARVPRLQWFRDVAALETPAGLPLREFPFAVYLITGNAGSGKSTCVQT
                     LNEVLDCVVTGATRIAAQNMYVKLSGAFLSRPINTIFHEFGFRGNHVQAQVGQHPYTL
                     ASNPASVEDLQRRDLTYYWEVIADITRRAVAAGGGAEDGRNEFRALATLESTMGLGEG
                     ALTRLAAFTHGALPAFTRSNIVVIDEAGLLGRHLLTAVVYCWWMTNALYHTPQYAAGL
                     RPVLVCVGSPTQTASLESTFEHQKLRCSIRQSENVLTYLICNRTLREYARLPRSWAIF
                     INNKRCVEHEFGNLMKALEYGLPLTEEHMQFVDRFVVPESFITNPANLPGWTRLFSSH
                     KEVSAYMARLHAHLKVAREGEFVVFSLPVLTFVSAKAFDEYRELTHQPALTVEKWIAA
                     NAGRITNYSQSQDQDAGRARLEVHSKQQLVVARNDVTYVLNSQIAVTTRLRKLVFGFS
                     GTFRSFEAVLRDDSFVKTQGEMLVEFAYQFLSRLIFGGLIHFYNFLLRPGLDAARRAR
                     AYERMGALTRELLSLGATGAGPATPGDQAASGGGESASAAPGAERAFDFKTLGARATD
                     AEFPEDDLDVVFAGLNEQELDLFYCHYTVEQPETTAAVHAQFGLLKRAFLGRYHILRE
                     LFGEAFEGAPFGTYVDNVTFRGCEVMIGSMRGGLMSVALQTDSYTLTGFTYARVYAFA
                     EEVRRRHAAGGLAELLEESPLPYVVLRDQHGFMSVVNTNISEFVESVESRELAMATNA
                     DYGISSMLAMTITRSQGLSLDKVAICFTPGNLRLNSAYVAMSRTTSSEFLRMNLNPLR
                     ERHERDDVISEHILSALRDPDVLIVY"
gene                 14819. .16870
                     /gene="UL6"
CDS                  14819. .16870
                     /gene="UL6"
                     /note="role in DNA cleavage/packaging"
                     /codon_start=1
                     /product="capsid protein"
                     /protein_id="AAP41424.1"
                     /db_xref="GI:30844248"
                     /translation="MAERHPRAARPPARPFDAAPRPPDDGWVQVHPTPTTMLFREALL
                     GELGYTEGQGVYNVVRSSEAVTRQLQAAVFGALLNAAAHRDLEADWRRHVAARRLEPE
                     RLVRRYGGAPDGDVAAVAERVYDTWRAALRASLLDFARGLADCFAQGGPGGPVGFSRY
                     IDWLMCLGLVPVMRRPREGEVTQRLEAFLSDHALPRDLATVAGAVERAAPGLRELARA
                     FDSARIADYDRVCLFYHPRRGEWRVRDPAGAQRGECVVLWPPLWTGDRLLFDSPMQRL
                     SREIVACRALREHARVCRLRNAASVKVLLGRKNDGDRGAAGAARAVSRALGEDDASKA
                     GSAASRLVRLIINMKGMRHIGDINDTVRAYLDEAGGHLIDSAAVDPALPGFGRATGRP
                     AAGGADPAAPRQQQLRQAFQTSVVNNINGMLEGYINNLFGTIERLRETNTDLATQLRD
                     RDIELRRARLGALDRQQRAADPGRAPEAEPAGQDDLRTDYDIIDVSKAMDDDTYVANS
                     FQHQYVPSYAQDLERLSRLWEHELLRCFKISRHTNNQGRETAIFYSSGAIALFVAPYF
                     ASVLRAPAMGALITGPNGVLGEEELWEAIFKKTRVQTYLTDLAALFVADVRHAARSRP
                     PAPEPADPDRRVRERPRGRSRSPVDAGRPGGRGPEDAWGGRGDHGHLDPRRGGDGGRR
                     P"
gene                 16821. .17711
                     /gene="UL7"
CDS                  16821. .17711
                     /gene="UL7"
                     /note="role in DNA cleavage/packaging"
                     /codon_start=1
                     /product="capsid protein"
                     /protein_id="AAP41425.1"
                     /db_xref="GI:30844249"
                     /translation="MATSTPAEEETAAAALRRAVAGDRDLVGVAEAISHQALLRMACE
                     VRQATRHTVRFTATGVVRVNVTPDHRLHFVLDGCPDDALIASDAYFARCRDQSAHRGF
                     AFAVVLAREDHVHSLCVPPLVLPHRLSVFWPEHLEDFELACLLMYLENCPRAHATASA
                     FVKISAWLAALGRRTSPFDRVRCLLLRSCQWMLNTLMFMVHVEPFDDQYVLPHWCMAR
                     YLLANNPPPILTALFRAGPSRVFRLPVQPRASLECVAYNPQGVLAAGWASEDFRAGLV
                     SWWSSDAPKTQTSALFYRLV"
gene                 complement(17886. .20171)
                     /gene="UL8"
CDS                  complement(17886. .20171)
```

```
                     /gene="UL8"
                     /codon_start=1
                     /product="component of helicase-primase complex"
                     /protein_id="AAP41426.1"
                     /db_xref="GI:30844250"
                     /translation="MGDAGLVWAEESICAITLYTAWLPPRTRDRLHVLLYLFCRDARG
                     DGAARFAEVSIASTELQAFYGPAEISAAGAVAAARAAASPAAAPLETLNDPVLWRALY
                     ACALVALEREVGRAALFAPARVAWDPRTGLVARIEGLPDGRSPAPRAAALDVNAQIAT
                     DPLALAAHVAARPDARLAWARLAAIQDTPQCASAASLTVRIDTGAARFAREYTTLTFP
                     PVRREGAVADVFEVREAVLRPRGHAQAVTARVLVPRGYEYFAVGAEGFSAPALIAMFR
                     QWYLTVHAGPGALPPIFAFLGPEFEPRGGAVDHFSVLGFPGWATLKVSGGAAAAPPAA
                     LAAHAELAGAWPAAGARALGPARAWAAVSAAAAERLPPALRRPAAAGLRLLDPPAVIG
                     PVCLARFRFPGLQPRLLAALYELGLRGDPRAADPRDPLGRGARRWRDLAGGEGAAEQA
                     PWPREVLALATDEVRREYPAVARTLDGIVSAVGVGLEKAAGAVDMAVCGDGGRVWGV
                     LNVDPRDAAGAGAAAAAAERARAALAASASEALAGWGLRLDAPPPLVLEGTYTHAVLW
                     SQTGAWFWNSDTDEDRLEGFPLRGPAYAAAAAVARRALRAVVAAGSEGGAGAEEALRI
```

Query Match 8.9%; Score 80.2; DB 13; Length 156789;
Best Local Similarity 44.6%; Pred. No. 22;
Matches 362; Conservative 0; Mismatches 443; Indels 6; Gaps 1;

```
Qy         52 GGAGATCGGCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGA 111
Db     154451 GGCGCGCAGCGCGGGCGGCAGGGCGTCGGGCCCGGCGCGCGCGGCGTCGCGGTCCAGCGC 154392

Qy        112 GGCGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCG 171
Db     154391 CAGGTAGACGGGCCGCAGCGGCGCGGCCAGCCCCCAGCGCGCGCAGGCGCGGTGCGAGTG 154332

Qy        172 CACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGAC 231
Db     154331 CGCCTCGCCGCGGCGGAAGTCCGGGGCGCCGGCGGCGGCGGCCTCGCGGGCGGGCGAGTC 154272

Qy        232 GCGGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGA 291
Db     154271 GGCGGCGCGGCCGTCGAGCGCGGGCAGCACGGCGCGGCGGTACTCGCGCGGCGGCAGGGG 154212

Qy        292 CGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACAC 351
Db     154211 CACGGGCGTGTCGGGCCCGAGGCGCGTGCGCGCGCGGTAGCGCACGTTGGCGCCGCGGCC 154152

Qy        352 CGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGC 411
Db     154151 CAGGCGCGGCGGCGGCGCGTCCGGGTAGAGGCGCGCGCAGGCGGCCTCGACGCGCGCGAA 154092

Qy        412 CTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAA 471
Db     154091 GACGCCGGGCCCGAAGACGCGGCCGGAGGCCAGCACGGCGCGGCGCAGGTCGCGCGCGGC 154032

Qy        472 GAAGCGCCG------GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCA 525
Db     154031 GGGCCAGCGGACGTCGCACTGCGCGGCGGGCAGCACGGCGCAGGGCAGGTAGACGTGGCG 153972

Qy        526 GGCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAG 585
Db     153971 GCGCGCGACGGCGGGCCCGTCGGCCGGCCACTCCTCGGGGGGCACGGCGTCCACGACGAC 153912

Qy        586 CCAGAGATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGT 645
Db     153911 GAGGCGGCGGCCGGCGGCGGCGGCGAGCAGGCCGAGGTACTCGACGGCGCCGGCGAAGGC 153852

Qy        646 CTTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGT 705
Db     153851 CAGGTCCCGCGTGGAGAGCAGGAGCACGCCCTGCGCGCCCAGCGCCGAGACGTCGGGCGG 153792

Qy        706 CCTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAG 765
Db     153791 GCCGGTCCAGTTGCCGGCCCAGGCGGCCGTGTCGGGCCCGCAGAGGCGGTTGCCCAGCGC 153732

Qy        766 CGGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGC 825
Db     153731 GGCGAGCAGGAAGGAGAGGCCGCCGCGCGCGTCGGTCCAGGCGGGCGGGGCGGGGGAGGC 153672
```

```
Qy        826 GAGGCCGGGGATGTCGCGGGGGCGGAGCTCG 856
Db     153671 GCCGCCGAGCTCCTCGCCGGGCAGCGGCGCG 153641
```

```
RESULT 13
AR151706/c
LOCUS       AR151706                1722 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 15 from patent US 6232106.
ACCESSION   AR151706
VERSION     AR151706.1  GI:15117756
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1722)
  AUTHORS   Jensen,S.E., Aidoo,K.A. and Paradkar,A.S.
  TITLE     DNA sequence encoding enzymes of clavulanic acid biosynthesis
  JOURNAL   Patent: US 6232106-A 15 15-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .1722
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Query Match 8.9%; Score 80; DB 6; Length 1722;
Best Local Similarity 45.1%; Pred. No. 1.2e+02;
Matches 339; Conservative 0; Mismatches 410; Indels 3; Gaps 1;

```
Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239
```

```
Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||      | |  | |  ||  |    |||||||    || ||  || |
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 14
AR352527/c
LOCUS       AR352527               1722 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 15 from patent US 6589775.
ACCESSION   AR352527
VERSION     AR352527.1  GI:33757735
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1722)
  AUTHORS   Jensen,S.E., Aidoo,K.A. and Paradkar,A.S.
  TITLE     DNA sequence encoding enzymes of clavulanic acid biosynthesis
  JOURNAL   Patent: US 6589775-A 15 08-JUL-2003;
            The Governors of the University of Alberta; Edmonton;
            EPX;
FEATURES             Location/Qualifiers
     source          1. .1722
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             8.9%;  Score 80;  DB 6;  Length 1722;
  Best Local Similarity   45.1%;  Pred. No. 1.2e+02;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |      |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    ||||  |      ||| |||  ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |   | ||| |      ||  || |    |  |||| || | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||          |||   | || || |  | ||| |||  || ||      ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||   || || | || ||| ||   |||   |||  |      |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||        || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||         || |  ||  |   || ||  |  |  | |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||   | |  | ||| ||   |  | ||   | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
                ||     |  |  |||  ||   | | ||   |  |||  |||  |      |
```

```
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |      |||    |||   |    |  |  | | |
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||      | |  | |  ||  |    |||||||    || ||  || |
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 15
AR151704/c
LOCUS       AR151704              11604 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 13 from patent US 6232106.
ACCESSION   AR151704
VERSION     AR151704.1  GI:15117754
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 11604)
  AUTHORS   Jensen,S.E., Aidoo,K.A. and Paradkar,A.S.
  TITLE     DNA sequence encoding enzymes of clavulanic acid biosynthesis
  JOURNAL   Patent: US 6232106-A 13 15-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .11604
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             8.9%;  Score 80;  DB 6;  Length 11604;
  Best Local Similarity   45.1%;  Pred. No. 59;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |      |  || | ||
Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    ||||  |      ||| |||  ||
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |   | ||| |      ||  || |    |  |||| || | |
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||          |||   | || || |  | ||| |||  || ||      ||
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||   || || | || ||| ||   |||   |||  |      |
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||        || | || || |  |  |||   | ||| | ||
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||         || |  ||  |   || ||  |  |  | |||||||  | |||
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||   | |  | ||| ||   |  | ||   | | ||| |  || ||
```

```
Db        601 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||      |  |  |||  ||   | | ||   |  |||  |||  |       |
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |     |||    |||   |    |   |  | |
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||      | |  | |  ||  |    |||||||    || ||  || |
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 330
```

Search completed: April 28, 2006, 02:39:44
Job time : 4995.51 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 185.89 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_76000_76900
Perfect score: 901
Sequence: 1 ggcgggggccttccggctgg.....cgggcgcggcccacgcgtcg 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	81.2	9.0	11604	3	US-09-385-040-13	Sequence 13, Appl
c	2	80	8.9	1722	3	US-09-385-028-15	Sequence 15, Appl
c	3	80	8.9	1722	3	US-09-726-614-15	Sequence 15, Appl
c	4	80	8.9	1722	3	US-09-385-040-15	Sequence 15, Appl
c	5	80	8.9	11604	3	US-09-385-028-13	Sequence 13, Appl
c	6	80	8.9	11604	3	US-09-726-614-13	Sequence 13, Appl
c	7	80	8.9	15079	3	US-09-385-028-1	Sequence 1, Appli
c	8	80	8.9	15079	3	US-09-726-614-1	Sequence 1, Appli
c	9	80	8.9	15120	3	US-09-385-040-1	Sequence 1, Appli
c	10	77.2	8.6	58857	3	US-09-477-962-1	Sequence 1, Appli
c	11	75.2	8.3	11220	3	US-09-105-537-32	Sequence 32, Appl
c	12	75.2	8.3	36778	3	US-09-105-537-5	Sequence 5, Appli
c	13	75.2	8.3	38506	3	US-09-320-878-19	Sequence 19, Appl
c	14	75.2	8.3	38506	3	US-09-141-908-1	Sequence 1, Appli
c	15	75.2	8.3	38506	3	US-09-657-440-19	Sequence 19, Appl
c	16	75.2	8.3	38506	3	US-09-793-708-19	Sequence 19, Appl
	17	73.4	8.1	11220	3	US-09-105-537-32	Sequence 32, Appl
	18	73.4	8.1	36778	3	US-09-105-537-5	Sequence 5, Appli
c	19	72.8	8.1	71989	3	US-09-443-501A-2	Sequence 2, Appli
c	20	72.8	8.1	71989	3	US-09-724-889A-2	Sequence 2, Appli
c	21	72.8	8.1	71989	3	US-09-724-876-2	Sequence 2, Appli
c	22	72.2	8.0	1320	3	US-09-724-797-85	Sequence 85, Appl
c	23	72.2	8.0	5802	3	US-10-152-886-84	Sequence 84, Appl
c	24	72.2	8.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
c	25	72	8.0	5877	3	US-10-152-886-94	Sequence 94, Appl
	26	71.8	8.0	3957	3	US-10-237-551-193	Sequence 193, App
	27	71.8	8.0	154746	3	US-09-827-688-8	Sequence 8, Appli
c	28	71.8	8.0	154746	3	US-09-827-688-8	Sequence 8, Appli
c	29	71.6	7.9	47981	3	US-09-679-279-1	Sequence 1, Appli
c	30	70.8	7.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
c	31	70.8	7.9	44377	2	US-08-804-198-1	Sequence 1, Appli
c	32	70.8	7.9	53799	3	US-10-042-665A-3	Sequence 3, Appli
	33	68.8	7.6	77536	3	US-09-410-551B-1	Sequence 1, Appli
	34	68.8	7.6	77536	3	US-09-940-316B-1	Sequence 1, Appli
	35	68.2	7.6	2721	9	5215881-2	Patent No. 5215881
	36	68.2	7.6	8438	2	US-07-945-283-1	Sequence 1, Appli
c	37	67.6	7.5	1926	3	US-09-249-585A-4	Sequence 4, Appli
c	38	67.6	7.5	1931	2	US-09-130-114-2	Sequence 2, Appli
c	39	66.2	7.3	987	3	US-09-902-540-5740	Sequence 5740, Ap
c	40	66.2	7.3	54484	3	US-09-902-540-1272	Sequence 1272, Ap
c	41	65	7.2	3157	2	US-08-939-002A-1	Sequence 1, Appli
c	42	64.6	7.2	1839	3	US-09-758-759-150	Sequence 150, App
c	43	64.6	7.2	109519	3	US-09-758-759-1	Sequence 1, Appli
c	44	64.2	7.1	1629	3	US-09-902-540-8462	Sequence 8462, Ap
c	45	64.2	7.1	1686	3	US-09-902-540-9105	Sequence 9105, Ap

ALIGNMENTS

RESULT 1
US-09-385-040-13/c
; Sequence 13, Application US/09385040
; Patent No. 6589775
; GENERAL INFORMATION:
; APPLICANT: Jensen, Susan E
; APPLICANT: Aidoo, Kwamena A
; APPLICANT: Paradkar, Ashish S
; TITLE OF INVENTION: DNA SEQUENCE ENCODING ENZYMES OF CLAVULANIC ACID
; TITLE OF INVENTION: BIOSYNTHESIS
; FILE REFERENCE: 09/385,040
; CURRENT APPLICATION NUMBER: US/09/385,040
; CURRENT FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: US 08/790,462
; PRIOR FILING DATE: 1997-01-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 11604
; TYPE: DNA
; ORGANISM: Streptomyces clavuligerus
US-09-385-040-13

Query Match 9.0%; Score 81.2; DB 3; Length 11604;
Best Local Similarity 45.1%; Pred. No. 1.2e-05;
Matches 342; Conservative 0; Mismatches 413; Indels 3; Gaps 1;

```
Qy      114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
            || ||    |   || || || ||    | ||| |   |  |    |  || | ||
Db     1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy      174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
            |   |||| ||  |   |   | ||| ||| |    |||| |       ||| ||| ||
Db     1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy      234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
             ||| ||| | |   |  | ||| |      ||   || |   |    |||| ||  | |
Db      961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy      291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
            |   ||||          |||   | || || |  | ||| |||  || ||     ||
Db      901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy      351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
            ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db      841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782
```

```
Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |     |
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
               || ||  | || |  ||      || | || || |  |  |||   | ||| | ||
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
                |||        || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||  | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        601 GGAGCTCCACGGCGTACAAGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
               ||     |  |  |||  ||   | | ||   |  |||  |||  |     |
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |      |||    |||   |   |   |  | ||
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||     | |  | |  ||  |   |||||||    || ||  || |
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCG 868
              ||| ||  ||| || ||   ||  ||  | ||  ||||
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCGATGGGG 324


RESULT 2
US-09-385-028-15/c
; Sequence 15, Application US/09385028
; Patent No. 6232106
;  GENERAL INFORMATION:
;    APPLICANT:  Susan E. Jensen
;    APPLICANT:  Kwamena A Aidoo
;    APPLICANT:  Ashish S. Paradkar
;    TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
;    TITLE OF INVENTION:  Acid Biosynthesis
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN,PPLC
;      STREET:  The Jenifer Buliding, 400 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/385,028
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/790,462
;      FILING DATE:  29-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D. Douglas Price
;      REGISTRATION NUMBER:  24,514
;      REFERENCE/DOCKET NUMBER:  1418/P57452US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202 638-6666
;      TELEFAX:  (202) 39305350
;      TELEX:  RCA 248593 IDEA UR
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-385-028-15

  Query Match             8.9%;  Score 80;  DB 3;  Length 1722;
  Best Local Similarity   45.1%;  Pred. No. 2.1e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |     |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |  |||| ||  |   |  | ||| ||| |    |||| |      ||| |||  ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |   |  | ||| |      ||   || |   |    |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |  ||||      |||   | || || |  | ||| |||  || ||        ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |     |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
               || ||  | || |  ||      || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
                |||        || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||  | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
               ||     |  |  |||  ||   | | ||   |  |||  |||  |     |
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |      |||    |||   |   |   |  | ||
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||     | |  | |  ||  |   |||||||    || ||  || |
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 3
US-09-726-614-15/c
; Sequence 15, Application US/09726614
; Patent No. 6514735
;  GENERAL INFORMATION:
;    APPLICANT:  Susan E. Jensen
;    APPLICANT:  Kwamena A Aidoo
;    APPLICANT:  Ashish S. Paradkar
```

```
;    TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6514735
;    TITLE OF INVENTION:  Acid Biosynthesis
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN,PPLC
;      STREET:  The Jenifer Buliding, 400 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/726,614
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/790,462
;      FILING DATE:  29-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D. Douglas Price
;      REGISTRATION NUMBER:  24,514
;      REFERENCE/DOCKET NUMBER:  1418/P57452US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202 638-6666
;      TELEFAX:  (202) 39305350
;      TELEX:  RCA 248593 IDEA UR
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-726-614-15

  Query Match             8.9%;  Score 80;  DB 3;  Length 1722;
  Best Local Similarity   45.1%;  Pred. No. 2.1e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |     |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    |||| |       ||| |||  ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               |||  ||| | |  |   | ||| |      ||   || |    |   |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||          |||   | || || |  | |||  |||  || ||      ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||  ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
```

```
              |||         || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |    ||||  | |  | ||| ||   |  |   | ||  | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||      |  | |||  ||   | | ||   |  |||  |||  |       |
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |         || |  |       |||     |||    |   |   |  | |
Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||      |  |  |  ||  |     |||||||     ||  ||  ||  |
Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              |||  ||  |||  ||  ||    ||  ||  |  ||
Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 4
US-09-385-040-15/c
; Sequence 15, Application US/09385040
; Patent No. 6589775
; GENERAL INFORMATION:
;  APPLICANT: Jensen, Susan E
;  APPLICANT:  Aidoo, Kwamena A
;  APPLICANT:  Paradkar, Ashish S
;  TITLE OF INVENTION: DNA SEQUENCE ENCODING ENZYMES OF CLAVULANIC ACID
;  TITLE OF INVENTION:  BIOSYNTHESIS
;  FILE REFERENCE: 09/385,040
;  CURRENT APPLICATION NUMBER: US/09/385,040
;  CURRENT FILING DATE:  1999-08-30
;  PRIOR APPLICATION NUMBER: US 08/790,462
;  PRIOR FILING DATE: 1997-01-29
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
;   LENGTH: 1722
;   TYPE: DNA
;   ORGANISM: Streptomyces clavuligerus
US-09-385-040-15

  Query Match             8.9%;  Score 80;  DB 3;  Length 1722;
  Best Local Similarity   45.1%;  Pred. No. 2.1e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |     |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    |||| |       ||| |||  ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               |||  ||| | |  |   | ||| |      ||   || |    |   |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||          |||   | || || |  | |||  |||  || ||      ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||  ||||||  |  |  | |  ||  ||  | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
```

```
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530

Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590

Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650

Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710

Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770

Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830

Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862

Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 5
US-09-385-028-13/c
; Sequence 13, Application US/09385028
; Patent No. 6232106
;  GENERAL INFORMATION:
;    APPLICANT:  Susan E. Jensen
;    APPLICANT:  Kwamena A Aidoo
;    APPLICANT:  Ashish S. Paradkar
;    TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
;    TITLE OF INVENTION:  Acid Biosynthesis
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN,PPLC
;      STREET:  The Jenifer Buliding, 400 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/385,028
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/790,462
;      FILING DATE:  29-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D. Douglas Price
;      REGISTRATION NUMBER:  24,514
;      REFERENCE/DOCKET NUMBER:  1418/P57452US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202 638-6666
;      TELEFAX:  (202) 39305350
;      TELEX:  RCA 248593 IDEA UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11604 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-385-028-13

  Query Match             8.9%;  Score 80;  DB 3;  Length 11604;
  Best Local Similarity   45.1%;  Pred. No. 1.9e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173

Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233

Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290

Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350

Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410

Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470

Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530

Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590

Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650

Db        601 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710

Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770

Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830

Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862

Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 330


RESULT 6
US-09-726-614-13/c
; Sequence 13, Application US/09726614
; Patent No. 6514735
;  GENERAL INFORMATION:
;    APPLICANT:  Susan E. Jensen
;    APPLICANT:  Kwamena A Aidoo
;    APPLICANT:  Ashish S. Paradkar
;    TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6514735
```

```
;    TITLE OF INVENTION:  Acid Biosynthesis
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN,PPLC
;      STREET:  The Jenifer Buliding, 400 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/726,614
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/790,462
;      FILING DATE:  29-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D. Douglas Price
;      REGISTRATION NUMBER:  24,514
;      REFERENCE/DOCKET NUMBER:  1418/P57452US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202 638-6666
;      TELEFAX:  (202) 39305350
;      TELEX:  RCA 248593 IDEA UR
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11604 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-726-614-13

  Query Match             8.9%;  Score 80;  DB 3;  Length 11604;
  Best Local Similarity   45.1%;  Pred. No. 1.9e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |  || || || ||    | ||| |   |  |      |  || | ||
Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    |||| |      ||| |||  ||
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               |||  ||| | |  |   | ||| |     ||   || |   |    |||| ||  | |
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |   ||||           |||   | || || |  | ||| |||  || ||     ||
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |     |
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | ||| |  ||       || | || || |  |  |||   | ||| | ||
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||          || |  ||  |   || ||  |  |  |  |||||||  | |||
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||  | | | ||| ||   |  |   | ||   | | ||| |  || ||
Db        601 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||      |  |  ||| ||  | | ||   | ||| |||      |
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |          || | |     |||   |||   |   |  |  | | |
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                ||      | | | | ||  |    |||||||    || ||  || |
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 330


RESULT 7
US-09-385-028-1/c
; Sequence 1, Application US/09385028
; Patent No. 6232106
;  GENERAL INFORMATION:
;    APPLICANT:  Susan E. Jensen
;    APPLICANT:  Kwamena A Aidoo
;    APPLICANT:  Ashish S. Paradkar
;    TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
;    TITLE OF INVENTION:  Acid Biosynthesis
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN,PPLC
;      STREET:  The Jenifer Buliding, 400 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/385,028
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/790,462
;      FILING DATE:  29-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D. Douglas Price
;      REGISTRATION NUMBER:  24,514
;      REFERENCE/DOCKET NUMBER:  1418/P57452US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202 638-6666
;      TELEFAX:  (202) 39305350
;      TELEX:  RCA 248593 IDEA UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces clavuligerus
US-09-385-028-1
```

```
  Query Match             8.9%;  Score 80;  DB 3;  Length 15079;
  Best Local Similarity   45.1%;  Pred. No. 1.9e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |     |  || | ||
Db       3113 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3054

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    |||| |       ||| ||| ||
Db       3053 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 2994

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               ||| ||| | |  |  | ||| |     ||  || |   |    |||| ||  | |
Db       2993 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2934

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||          |||   | || || |  | ||| |||  || ||     ||
Db       2933 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2874

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db       2873 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2814

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |   || |||   || || | || ||| ||   |||   ||| |     |
Db       2813 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2754

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
               || || |  | || |  ||      || | || || |  |  |||   | ||| | ||
Db       2753 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2694

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||         || |  ||  |   || ||  |  |  |  |||||||  | |||
Db       2693 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2634

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |   ||||  | |  | ||| ||   |  |  | ||  | | ||| |  || ||
Db       2633 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2574

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  | ||| ||   | | ||   | ||| |||    |  |       |
Db       2573 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2514

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |     |||    |||   |   |  |  | | |
Db       2513 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2454

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
                 ||     | |  | |  ||  |    |||||||    || ||  || |
Db       2453 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2394

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db       2393 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2362


RESULT 8
US-09-726-614-1/c
; Sequence 1, Application US/09726614
; Patent No. 6514735
;  GENERAL INFORMATION:
;    APPLICANT:  Susan E. Jensen
;    APPLICANT:  Kwamena A Aidoo
;    APPLICANT:  Ashish S. Paradkar
;    TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6514735
;    TITLE OF INVENTION:  Acid Biosynthesis
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN,PPLC
;      STREET:  The Jenifer Buliding, 400 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/726,614
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/790,462
;      FILING DATE:  29-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D. Douglas Price
;      REGISTRATION NUMBER:  24,514
;      REFERENCE/DOCKET NUMBER:  1418/P57452US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202 638-6666
;      TELEFAX:  (202) 39305350
;      TELEX:  RCA 248593 IDEA UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces clavuligerus
US-09-726-614-1

  Query Match             8.9%;  Score 80;  DB 3;  Length 15079;
  Best Local Similarity   45.1%;  Pred. No. 1.9e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    |  ||| |    |  |     |  || | ||
Db       3113 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3054

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |   | ||| ||| |    |||| |       ||| ||| ||
Db       3053 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 2994

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
               ||| ||| | |  |  | ||| |     ||  || |   |    |||| ||  | |
Db       2993 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2934

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||          |||   | || || |  | ||| |||  || ||     ||
Db       2933 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2874

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db       2873 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2814

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |   || |||   || || | || ||| ||   |||   ||| |     |
Db       2813 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2754

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
               || || |  | || |  ||      || | || || |  |  |||   | ||| | ||
Db       2753 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2694

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||         || |  ||  |   || ||  |  |  |  |||||||  | |||
Db       2693 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2634

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |   ||||  | |  | ||| ||   |  |  | ||  | | ||| |  || ||
```

```
Db       2633 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2574

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710

Db       2573 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2514

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770

Db       2513 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2454

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830

Db       2453 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2394

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862

Db       2393 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2362


RESULT 9
US-09-385-040-1/c
; Sequence 1, Application US/09385040
; Patent No. 6589775
; GENERAL INFORMATION:
;  APPLICANT: Jensen, Susan E
;  APPLICANT:  Aidoo, Kwamena A
;  APPLICANT:  Paradkar, Ashish S
;  TITLE OF INVENTION: DNA SEQUENCE ENCODING ENZYMES OF CLAVULANIC ACID
;  TITLE OF INVENTION:  BIOSYNTHESIS
;  FILE REFERENCE: 09/385,040
;  CURRENT APPLICATION NUMBER: US/09/385,040
;  CURRENT FILING DATE:  1999-08-30
;  PRIOR APPLICATION NUMBER: US 08/790,462
;  PRIOR FILING DATE: 1997-01-29
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 15120
;   TYPE: DNA
;   ORGANISM: Streptomyces clavuligerus
US-09-385-040-1

  Query Match             8.9%;  Score 80;  DB 3;  Length 15120;
  Best Local Similarity   45.1%;  Pred. No. 1.9e-05;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173

Db       3154 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3095

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233

Db       3094 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 3035

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290

Db       3034 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2975

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350

Db       2974 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2915

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410

Db       2914 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2855

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470

Db       2854 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2795

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530

Db       2794 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2735


Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590

Db       2734 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2675

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650

Db       2674 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2615

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710

Db       2614 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2555

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770

Db       2554 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2495

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830

Db       2494 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2435

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862

Db       2434 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2403


RESULT 10
US-09-477-962-1/c
; Sequence 1, Application US/09477962
; Patent No. 6927286
; GENERAL INFORMATION:
;  APPLICANT: SHEN, BEN
;  APPLICANT:  DU, LIANGCHENG
;  APPLICANT:  SANCHEZ, CESAR
;  APPLICANT:  CHEN, MEI
;  APPLICANT:  EDWARDS, DANIEL J.
;  TITLE OF INVENTION: BLEOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES
;  FILE REFERENCE: 407T-895820US
;  CURRENT APPLICATION NUMBER: US/09/477,962
;  CURRENT FILING DATE:  2000-01-05
;  PRIOR APPLICATION NUMBER: 60/115,435
;  PRIOR FILING DATE: 1999-01-06
;  PRIOR APPLICATION NUMBER: 60/118,848
;  PRIOR FILING DATE: 1999-02-05
;  NUMBER OF SEQ ID NOS: 133
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1
;   LENGTH: 58857
;   TYPE: DNA
;   ORGANISM: Streptomyces verticillus
US-09-477-962-1

  Query Match             8.6%;  Score 77.2;  DB 3;  Length 58857;
  Best Local Similarity   45.3%;  Pred. No. 5.6e-05;
  Matches  406;  Conservative    0;  Mismatches  478;  Indels   12;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60

Db      28169 GACGGGCCGCTCCGGGCCGGCGCCCTCCCGGCGGAGCCGGGCGGCGAGGGTGGTGGCCGC 28110

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120

Db      28109 GCGGTGCAGGGCTCCGTGGCTGATGTGGTGGGCGGTGCCGTCCGGGGCGAGTGCGACGAC 28050

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180

Db      28049 CGCCGTCCGTGCGGGATCGCGGCGCGCCGCCGCGGCCACGAGCCGGTGCAGCGGGAGGGC 27990

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240

Db      27989 GGGCCGGGCCGCCGACGGGCCGCGGCGCGGCCCGGTCTCCCGGCCGCCGGTGAGCGGCAG 27930

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGC---CGGACCGACGCCTC 297
```

```
Db      27929 CGTCTCCACGGGCGCGCCCGGGTCCTCGACGGCCGCCCGCAGCAGGTGGACGAACGCCTC 27870

Qy        298 GCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGG 357

Db      27869 GCTCAGCTCCGCGACCGTGGCCTCGGTGAACAGGTCGGTGCGGTACTCCCACACCCCCGT 27810

Qy        358 CGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCG 417

Db      27809 CAGCCCGTCCCCGAGCCGCGCCATGCTCAGCGAGA------GGTCGAGCTGCGACCAGCG 27756

Qy        418 GCGCGCGCCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCG 477

Db      27755 GCGCGGCAACGCGAGTGCCTCCAGGTCGAGTCCGCCGACGCGCAGCCGGCCGCCGACGCC 27696

Qy        478 CCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGAC 537

Db      27695 CAGGGCGAGCGCGCGGTACCCGTCCGCCTCGTCGCCGTACTCGCGCTGGAGCACGAACAG 27636

Qy        538 GCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCG 597

Db      27635 GCTCTGCAGGAGCGGCGCGCGGCCGGGCGTGCGCGCGGGGGCGAGACGGGAGACGAGCGG 27576

Qy        598 CACGA---GCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCAT 654

Db      27575 CCCGAAGGGGTGGCGGCCGTGCGCGACCGCGTCCAGCACGGTCCGCCGGGTGCGGCGCAG 27516

Qy        655 GAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTG 714

Db      27515 CAGTTCCGTGAAGGTCTCCCCCGGCTCCCGTACGGAGCGCAGCGGCAGCGGGTTGACCAG 27456

Qy        715 GGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAA 774

Db      27455 GTAGCCGACGGCGCCGGCCGCTTCGGCGGTGTCGCGGCGGGCGAGGAGGGTGCCCACGAC 27396

Qy        775 GGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGG 834

Db      27395 GAGGTCGTCCTGCCCGGTCAGGCGGTGGAGCAGTAGCTGGTGGGCGGCGAGGAGGGTGGT 27336

Qy        835 GATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGG 890

Db      27335 GAAGAGGGTGCAGTGCCGTTCCCGGGAGAGGGCGGTCAGCCGGGCGGTGAGGTCCG 27280


RESULT 11
US-09-105-537-32/c
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             8.3%;  Score 75.2;  DB 3;  Length 11220;
  Best Local Similarity   45.2%;  Pred. No. 0.00013;
  Matches  409;  Conservative    0;  Mismatches  478;  Indels   17;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60

Db      10120 GGAGGGTGCCGTCGGAGAGGACGCCTGCCGTGTGGACGACCGCGGTGAGCGGGTGTTCGG 10061


Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120

Db      10060 CGGGGATCGAGTCGAGTACGGCGGTGA--GGGCTTCGCGGTCGGCGACGTCGCACGCGGC 10003

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180

Db      10002 CACCGAGACGTCGGCTCCCAGGGCCTCCAGCTCGTGCACGAGCTCGCCGGCGCCCGGGGC 9943

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240

Db       9942 GTCCGTGCCCCGCCGGCTCACGAGCAGCAGGCGTCGTACGCCCCACTCGCCCACCACGTG 9883

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300

Db       9882 CCGGGCCACGATGCCCCCCAGCGCACCGGTGCCGCCGGTCAGCAGGACCGTACCCTCCGG 9823

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360

Db       9822 GTCGAGGCCCGACGGCATCGTGAGGACGACCTTGCCCGTGTGGC------------GGGC 9775

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420

Db       9774 CTGGCTGACGTGCCGGAAGGCGTCGCGGGCCCGGCGCACGTCCCAGGTCGTGACGGGCAG 9715

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480

Db       9714 GTGCCGGAGCACCCCGTCCTCGAAGAGGGCGATGACCTCGGCGAGCATCTCGCCGATCCG 9655

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540

Db       9654 CTCCGGCCCGGCCTCGCCCAGGTCGAAGGCGCGGTAGCCGACACCGGGGTGGTCGGCGGC 9595

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600

Db       9594 GACCCGCTCCGCGTCGCGGACGTCGGTCTTCCCCATCTCCACGAACCGGCCGCCCGGCCC 9535

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660

Db       9534 GAGCAGGCGCAGCGAGGCGTCGACGAACTCGCGGGCGAGCGAGTTCAGTACGACGTCCAT 9475

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAA---GAGATCAGGCGTCCTGCGCTGGGT 717

Db       9474 GCCCGCCCCGCCGGAAGCGGCACGGAACGCGGACTCGAAGTCCAGGGTGCGGGAGGAGGC 9415

Qy        718 CGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGT 777

Db       9414 GATGTGCGCGTCGTCCAGGCCGAGCGCGCGCAGGGCGTCCCACTTCCCGTGACTCGCCGT 9355

Qy        778 GCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGAT 837

Db       9354 GCCGTGGACCTCCACGCCCCAGTGCCGGGCGAGCTGCACGGCGGCCATGCCCACGCCACC 9295

Qy        838 GTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGC 897

Db       9294 GGCGGCGGAGTGGACCAGGAGGCGCTCGCCGGGCTTGACGTCCGCCAGGTCGCGCAGGGC 9235

Qy        898 GTCG 901

Db       9234 GTAG 9231


RESULT 12
US-09-105-537-5/c
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
```

```
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             8.3%;  Score 75.2;  DB 3;  Length 36778;
  Best Local Similarity   45.2%;  Pred. No. 0.00012;
  Matches  409;  Conservative    0;  Mismatches  478;  Indels   17;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              || ||| ||| || |   ||| |  | |   | |  ||| |       |  || | ||||
Db      25807 GGAGGGTGCCGTCGGAGAGGACGCCTGCCGTGTGGACGACCGCGGTGAGCGGGTGTTCGG 25748

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              || ||  ||    |    ||  ||  |  ||||  |||||   |      |    ||
Db      25747 CGGGGATCGAGTCGAGTACGGCGGTGA--GGGCTTCGCGGTCGGCGACGTCGCACGCGGC 25690

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              | ||  |  |  |||   | || | || || ||| |  || |      ||    |   |
Db      25689 CACCGAGACGTCGGCTCCCAGGGCCTCCAGCTCGTGCACGAGCTCGCCGGCGCCCGGGGC 25630

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              |    |  ||||  | | | | | |||  |        | | |   ||| |    ||| |
Db      25629 GTCCGTGCCCCGCCGGCTCACGAGCAGCAGGCGTCGTACGCCCCACTCGCCCACCACGTG 25570

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              |  |||| ||  | |  | ||| |  |  ||||| |  ||||| ||||||   ||||  |
Db      25569 CCGGGCCACGATGCCCCCCAGCGCACCGGTGCCGCCGGTCAGCAGGACCGTACCCTCCGG 25510

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              ||| ||  |    ||  || | |   ||     ||| || ||              | |
Db      25509 GTCGAGGCCCGACGGCATCGTGAGGACGACCTTGCCCGTGTGGC-----------GGGC 25462

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              | |   | |||   ||  |  ||    ||||  | || |||||| ||||   || |  |
Db      25461 CTGGCTGACGTGCCGGAAGGCGTCGCGGGCCCGGCGCACGTCCCAGGTCGTGACGGGCAG 25402

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
                |      ||    | | |   | |||  |  |  |  | ||     |  |  |||
Db      25401 GTGCCGGAGCACCCCGTCCTCGAAGAGGGCGATGACCTCGGCGAGCATCTCGCCGATCCG 25342

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
              | ||||  |    |  ||   ||||  |  |  |    |  ||  | ||||  ||
Db      25341 CTCCGGCCCGGCCTCGCCCAGGTCGAAGGCGCGGTAGCCGACACCGGGGTGGTCGGCGGC 25282

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              | |  |       | ||  ||      ||||| |||| || |      ||  |
Db      25281 GACCCGCTCCGCGTCGCGGACGTCGGTCTTCCCCATCTCCACGAACCGGCCGCCCGGCCC 25222

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660
              ||||   || | |   || ||  |  ||| |||| |  |  | ||      | |  ||
Db      25221 GAGCAGGCGCAGCGAGGCGTCGACGAACTCGCGGGCGAGCGAGTTCAGTACGACGTCCAT 25162

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAA---GAGATCAGGCGTCCTGCGCTGGGT 717
                  |  || | |  || ||| | |||  |   ||  || | | |  |      ||
Db      25161 GCCCGCCCCGCCGGAAGCGGCACGGAACGCGGACTCGAAGTCCAGGGTGCGGGAGGAGGC 25102

Qy        718 CGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGT 777
              |  | ||  ||||     |  |  |      | |      |       |       ||
Db      25101 GATGTGCGCGTCGTCCAGGCCGAGCGCGCGCAGGGCGTCCCACTTCCCGTGACTCGCCGT 25042

Qy        778 GCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGAT 837
              ||||||  |  | |   | | |    | |||||||||| ||||  |    |    ||
Db      25041 GCCGTGGACCTCCACGCCCCAGTGCCGGGCGAGCTGCACGGCGGCCATGCCCACGCCACC 24982

Qy        838 GTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGC 897
              | ||  || | ||| |  | ||    || || ||  |  | |||  |  ||  |   ||
Db      24981 GGCGGCGGAGTGGACCAGGAGGCGCTCGCCGGGCTTGACGTCCGCCAGGTCGCGCAGGGC 24922

Qy        898 GTCG 901
                || |
Db      24921 GTAG 24918


RESULT 13
US-09-320-878-19/c
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             8.3%;  Score 75.2;  DB 3;  Length 38506;
  Best Local Similarity   45.2%;  Pred. No. 0.00012;
  Matches  409;  Conservative    0;  Mismatches  478;  Indels   17;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              || ||| ||| || |   ||| |  | |   | |  ||| |       |  || | ||||
Db      23949 GGAGGGTGCCGTCGGAGAGGACGCCTGCCGTGTGGACGACCGCGGTGAGCGGGTGTTCGG 23890

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              || ||  ||    |    ||  ||  |  ||||  |||||   |      |    ||
Db      23889 CGGGGATCGAGTCGAGTACGGCGGTGA--GGGCTTCGCGGTCGGCGACGTCGCACGCGGC 23832

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              | ||  |  |  |||   | || | || || ||| |  || |      ||    |   |
Db      23831 CACCGAGACGTCGGCTCCCAGGGCCTCCAGCTCGTGCACGAGCTCGCCGGCGCCCGGGGC 23772

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              |    |  ||||  | | | | | |||  |        | | |   ||| |    ||| |
Db      23771 GTCCGTGCCCCGCCGGCTCACGAGCAGCAGGCGTCGTACGCCCCACTCGCCCACCACGTG 23712

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              |  |||| ||  | |  | ||| |  |  ||||| |  ||||| ||||||   ||||  |
Db      23711 CCGGGCCACGATGCCCCCCAGCGCACCGGTGCCGCCGGTCAGCAGGACCGTACCCTCCGG 23652

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              ||| ||  |    ||  || | |   ||     ||| || ||              | |
Db      23651 GTCGAGGCCCGACGGCATCGTGAGGACGACCTTGCCCGTGTGGC-----------GGGC 23604

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              | |   | |||   ||  |  ||    ||||  | || |||||| ||||   || |  |
Db      23603 CTGGCTGACGTGCCGGAAGGCGTCGCGGGCCCGGCGCACGTCCCAGGTCGTGACGGGCAG 23544

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
                |      ||    | | |   | |||  |  |  |  | ||     |  |  |||
```

```
Db      23543 GTGCCGGAGCACCCCGTCCTCGAAGAGGGCGATGACCTCGGCGAGCATCTCGCCGATCCG 23484

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540

Db      23483 CTCCGGCCCGGCCTCGCCCAGGTCGAAGGCGCGGTAGCCGACACCGGGGTGGTCGGCGGC 23424

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600

Db      23423 GACCCGCTCCGCGTCGCGGACGTCGGTCTTCCCCATCTCCACGAACCGGCCGCCCGGCCC 23364

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660

Db      23363 GAGCAGGCGCAGCGAGGCGTCGACGAACTCGCGGGCGAGCGAGTTCAGTACGACGTCCAT 23304

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAA---GAGATCAGGCGTCCTGCGCTGGGT 717

Db      23303 GCCCGCCCCGCCGGAAGCGGCACGGAACGCGGACTCGAAGTCCAGGGTGCGGGAGGAGGC 23244

Qy        718 CGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGT 777

Db      23243 GATGTGCGCGTCGTCCAGGCCGAGCGCGCGCAGGGCGTCCCACTTCCCGTGACTCGCCGT 23184

Qy        778 GCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGAT 837

Db      23183 GCCGTGGACCTCCACGCCCCAGTGCCGGGCGAGCTGCACGGCGGCCATGCCCACGCCACC 23124

Qy        838 GTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGC 897

Db      23123 GGCGGCGGAGTGGACCAGGAGGCGCTCGCCGGGCTTGACGTCCGCCAGGTCGCGCAGGGC 23064

Qy        898 GTCG 901

Db      23063 GTAG 23060
```

RESULT 14
US-09-141-908-1/c
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1

Query Match 8.3%; Score 75.2; DB 3; Length 38506;
Best Local Similarity 45.2%; Pred. No. 0.00012;
Matches 409; Conservative 0; Mismatches 478; Indels 17; Gaps 3;

```
Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60

Db      23949 GGAGGGTGCCGTCGGAGAGGACGCCTGCCGTGTGGACGACCGCGGTGAGCGGGTGTTCGG 23890

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120

Db      23889 CGGGGATCGAGTCGAGTACGGCGGTGA--GGGCTTCGCGGTCGGCGACGTCGCACGCGGC 23832

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180

Db      23831 CACCGAGACGTCGGCTCCCAGGGCCTCCAGCTCGTGCACGAGCTCGCCGGCGCCCGGGGC 23772

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240

Db      23771 GTCCGTGCCCCGCCGGCTCACGAGCAGCAGGCGTCGTACGCCCCACTCGCCCACCACGTG 23712

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300

Db      23711 CCGGGCCACGATGCCCCCCAGCGCACCGGTGCCGCCGGTCAGCAGGACCGTACCCTCCGG 23652

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360

Db      23651 GTCGAGGCCCGACGGCATCGTGAGGACGACCTTGCCCGTGTGGC------------GGGC 23604

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420

Db      23603 CTGGCTGACGTGCCGGAAGGCGTCGCGGGCCCGGCGCACGTCCCAGGTCGTGACGGGCAG 23544

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480

Db      23543 GTGCCGGAGCACCCCGTCCTCGAAGAGGGCGATGACCTCGGCGAGCATCTCGCCGATCCG 23484

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540

Db      23483 CTCCGGCCCGGCCTCGCCCAGGTCGAAGGCGCGGTAGCCGACACCGGGGTGGTCGGCGGC 23424

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600

Db      23423 GACCCGCTCCGCGTCGCGGACGTCGGTCTTCCCCATCTCCACGAACCGGCCGCCCGGCCC 23364

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660

Db      23363 GAGCAGGCGCAGCGAGGCGTCGACGAACTCGCGGGCGAGCGAGTTCAGTACGACGTCCAT 23304

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAA---GAGATCAGGCGTCCTGCGCTGGGT 717

Db      23303 GCCCGCCCCGCCGGAAGCGGCACGGAACGCGGACTCGAAGTCCAGGGTGCGGGAGGAGGC 23244

Qy        718 CGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGT 777

Db      23243 GATGTGCGCGTCGTCCAGGCCGAGCGCGCGCAGGGCGTCCCACTTCCCGTGACTCGCCGT 23184

Qy        778 GCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGAT 837

Db      23183 GCCGTGGACCTCCACGCCCCAGTGCCGGGCGAGCTGCACGGCGGCCATGCCCACGCCACC 23124

Qy        838 GTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGC 897

Db      23123 GGCGGCGGAGTGGACCAGGAGGCGCTCGCCGGGCTTGACGTCCGCCAGGTCGCGCAGGGC 23064

Qy        898 GTCG 901

Db      23063 GTAG 23060
```

RESULT 15
US-09-657-440-19/c
; Sequence 19, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

```
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/657,440
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: CIP OF 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-657-440-19

  Query Match             8.3%;  Score 75.2;  DB 3;  Length 38506;
  Best Local Similarity   45.2%;  Pred. No. 0.00012;
  Matches  409;  Conservative    0;  Mismatches  478;  Indels   17;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              || ||| ||| || |   ||| |  | |   | |  ||| |       |  || | ||||
Db      23949 GGAGGGTGCCGTCGGAGAGGACGCCTGCCGTGTGGACGACCGCGGTGAGCGGGTGTTCGG 23890

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              || ||  ||    |   ||  ||  |  ||||  |||||   |       |    ||
Db      23889 CGGGGATCGAGTCGAGTACGGCGGTGA--GGGCTTCGCGGTCGGCGACGTCGCACGCGGC 23832

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              | ||  |  |  |||   | || | || || ||| |  || |      ||     |   |
Db      23831 CACCGAGACGTCGGCTCCCAGGGCCTCCAGCTCGTGCACGAGCTCGCCGGCGCCCGGGGC 23772

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              |      | ||||  | | | | | ||| |       | | |   ||| |    ||| |
Db      23771 GTCCGTGCCCCGCCGGCTCACGAGCAGCAGGCGTCGTACGCCCCACTCGCCCACCACGTG 23712

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              |  |||| ||  | |  | ||| |  |  ||||| |  ||||| ||||||  |||| |
Db      23711 CCGGGCCACGATGCCCCCCAGCGCACCGGTGCCGCCGGTCAGCAGGACCGTACCCTCCGG 23652

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              |||| ||  |    ||  || | |   ||     ||| || ||             | |
Db      23651 GTCGAGGCCCGACGGCATCGTGAGGACGACCTTGCCCGTGTGGC-----------GGGC 23604

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              | |   | |||   ||  |  ||    ||||  | || |||||| ||||   || |  |
Db      23603 CTGGCTGACGTGCCGGAAGGCGTCGCGGGCCCGGCGCACGTCCCAGGTCGTGACGGGCAG 23544

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
              |       ||    | | |    | |||  |  |  |  | ||     |  |  |||
Db      23543 GTGCCGGAGCACCCCGTCCTCGAAGAGGGCGATGACCTCGGCGAGCATCTCGCCGATCCG 23484

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
              | ||||| |   | ||   ||||  |  | |    |   | ||  | ||||  ||
Db      23483 CTCCGGCCCGGCCTCGCCCAGGTCGAAGGCGCGGTAGCCGACACCGGGGTGGTCGGCGGC 23424

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              | |      |     | ||  ||     | ||||| |||| || |       || |
Db      23423 GACCCGCTCCGCGTCGCGGACGTCGGTCTTCCCCATCTCCACGAACCGGCCGCCCGGCCC 23364

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660
              ||||   || | |   || ||  |   ||| |||| |  |  | ||      | |  ||
Db      23363 GAGCAGGCGCAGCGAGGCGTCGACGAACTCGCGGGCGAGCGAGTTCAGTACGACGTCCAT 23304

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAA---GAGATCAGGCGTCCTGCGCTGGGT 717
                  |  || | |  || ||| | ||| |   ||  ||  | || | |      ||
Db      23303 GCCCGCCCCGCCGGAAGCGGCACGGAACGCGGACTCGAAGTCCAGGGTGCGGGAGGAGGC 23244

Qy        718 CGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGT 777
              |  | ||  ||||       | |  |     | |     |       |      ||
Db      23243 GATGTGCGCGTCGTCCAGGCCGAGCGCGCGCAGGGCGTCCCACTTCCCGTGACTCGCCGT 23184

Qy        778 GCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGAT 837
              ||||||  |  | |   | | |    | ||||||||| ||||  |   |    ||

Db      23183 GCCGTGGACCTCCACGCCCCAGTGCCGGGCGAGCTGCACGGCGGCCATGCCCACGCCACC 23124

Qy        838 GTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGC 897
              | ||  || | ||| |  | ||     || || ||  |  |  |||  |  ||  |   ||
Db      23123 GGCGGCGGAGTGGACCAGGAGGCGCTCGCCGGGCTTGACGTCCGCCAGGTCGCGCAGGGC 23064

Qy        898 GTCG 901
              || |
Db      23063 GTAG 23060


Search completed: April 27, 2006, 12:42:41
Job time : 187.89 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 903.317 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_76000_76900
Perfect score: 901
Sequence: 1 ggcgggggccttccggctgg.....cgggcgcggcccacgcgtcg 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	901	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
c	2	90.4	10.0	1644	7	US-10-437-963-73289	Sequence 73289, A
c	3	82.4	9.1	1431	6	US-10-156-761-6757	Sequence 6757, Ap
c	4	82.4	9.1	9025608	6	US-10-156-761-1	Sequence 1, Appli
c	5	81.2	9.0	11604	6	US-10-458-201-13	Sequence 13, Appl
	6	80.2	8.9	3024	9	US-10-939-036-35	Sequence 35, Appl
c	7	80	8.9	1722	6	US-10-458-201-15	Sequence 15, Appl
c	8	80	8.9	15120	6	US-10-458-201-1	Sequence 1, Appli
c	9	80	8.9	29870	9	US-10-494-495-1	Sequence 1, Appli
c	10	77.8	8.6	5223	7	US-10-282-122A-14103	Sequence 14103, A
c	11	77.4	8.6	1431	6	US-10-084-846A-45	Sequence 45, Appl
c	12	77.4	8.6	59816	6	US-10-084-846A-1	Sequence 1, Appli
	13	77.4	8.6	59816	6	US-10-084-846A-2	Sequence 2, Appli
c	14	77	8.5	1491	6	US-10-156-761-1868	Sequence 1868, Ap
c	15	75.2	8.3	11220	3	US-09-861-289-32	Sequence 32, Appl
c	16	75.2	8.3	11220	3	US-09-860-846-32	Sequence 32, Appl
c	17	75.2	8.3	11220	3	US-09-988-384B-32	Sequence 32, Appl
c	18	75.2	8.3	11220	3	US-09-836-821-32	Sequence 32, Appl
c	19	75.2	8.3	11220	6	US-10-271-889-32	Sequence 32, Appl
c	20	75.2	8.3	36778	3	US-09-861-289-5	Sequence 5, Appli
c	21	75.2	8.3	36778	3	US-09-860-846-5	Sequence 5, Appli
c	22	75.2	8.3	36778	3	US-09-836-821-5	Sequence 5, Appli
c	23	75.2	8.3	36778	6	US-10-271-889-48	Sequence 48, Appl
c	24	75.2	8.3	37948	3	US-09-988-384B-5	Sequence 5, Appli
c	25	75.2	8.3	38506	3	US-09-793-708-19	Sequence 19, Appl
c	26	75.2	8.3	38506	6	US-10-201-365-1	Sequence 1, Appli
c	27	75.2	8.3	38506	6	US-10-160-539-19	Sequence 19, Appl
c	28	75.2	8.3	38506	8	US-10-468-828-19	Sequence 19, Appl
c	29	75.2	8.3	38506	9	US-10-846-335-19	Sequence 19, Appl
c	30	74.8	8.3	2118	7	US-10-282-122A-14081	Sequence 14081, A
c	31	74.8	8.3	22533	9	US-10-760-493-36	Sequence 36, Appl
c	32	74.8	8.3	164051	9	US-10-760-493-18	Sequence 18, Appl
c	33	74.2	8.2	1722	8	US-10-762-107-81	Sequence 81, Appl
c	34	74.2	8.2	9762	8	US-10-762-107-73	Sequence 73, Appl
c	35	73.4	8.1	1843	7	US-10-437-963-22190	Sequence 22190, A
	36	73.4	8.1	11220	3	US-09-861-289-32	Sequence 32, Appl
	37	73.4	8.1	11220	3	US-09-860-846-32	Sequence 32, Appl
	38	73.4	8.1	11220	3	US-09-988-384B-32	Sequence 32, Appl
	39	73.4	8.1	11220	3	US-09-836-821-32	Sequence 32, Appl
	40	73.4	8.1	11220	6	US-10-271-889-32	Sequence 32, Appl
	41	73.4	8.1	36778	3	US-09-861-289-5	Sequence 5, Appli
	42	73.4	8.1	36778	3	US-09-860-846-5	Sequence 5, Appli
	43	73.4	8.1	36778	3	US-09-836-821-5	Sequence 5, Appli
	44	73.4	8.1	36778	6	US-10-271-889-48	Sequence 48, Appl
	45	73.4	8.1	37948	3	US-09-988-384B-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 901; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 9.3e-214;
Matches 901; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76000 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 76059

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76060 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 76119

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76120 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 76179
```

```
Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCCGCAGATCGACGCGGACGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76180 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCCGCAGATCGACGCGGACGAG 76239

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76240 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 76299

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76300 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 76359

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76360 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 76419

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76420 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 76479

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76480 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 76539

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76540 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 76599

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76600 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 76659

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76660 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGC 76719

Qy        721 GGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76720 GGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCC 76779

Qy        781 GTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76780 GTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTC 76839

Qy        841 GCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76840 GCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTC 76899

Qy        901 G 901
              |
Db      76900 G 76900
```

RESULT 2
US-10-437-963-73289/c
; Sequence 73289, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 73289
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_73585C.1
US-10-437-963-73289

Query Match 10.0%; Score 90.4; DB 7; Length 1644;
Best Local Similarity 45.1%; Pred. No. 5.8e-13;
Matches 404; Conservative 0; Mismatches 471; Indels 21; Gaps 1;

```
Qy          3 CGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCG 62
              | | || | |||   |||     |   |   |||     |  || ||  |      ||| |
Db       1102 CCGAGGTCATCCACTTGGGCTCGCCGGTGAAGCGGACGAGGAGCTCCGCCATCGCCGGGG 1043

Qy         63 CGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCG 122
              ||   |   | | | | ||  |  |     |  ||||  | |   | || ||||||
Db       1042 CGTCGCCGAGGAGGCCGGAGAGCCTCCAGTTCCGGACGAGCACCAGGCGGTCGTCGAGCG 983

Qy        123 CCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGA 182
              |   | ||||  ||  ||| ||  | |   |  ||||  ||   |  || ||  | |
Db        982 CGTAGGTGAGGCTGAGGCGGGCGAGGCGCGGGGCGCCGCCGAAGGCCAGCGCGACGGGGC 923

Qy        183 TGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCA 242
              | |   | ||| | |     ||||| | | || | ||||| |||||  |  ||
Db        922 CGTCCACGCACGCCACGCGCTCCAGCGCCGGCGCCGCGCGCAGGTCGACGGTGTGGAAGC 863

Qy        243 CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGAT 302
              ||    || |  |||||||||          |   | ||     ||||  |  | ||  |
Db        862 CGCAGCCGTCCACGACGAGCTCGCGGAGCTCCGGCATCCCGTCGACCACCAGCGCGGCGT 803

Qy        303 CGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCG 362
              |   | |||  |   ||| |     | ||||  |  | |     ||||    ||
Db        802 ACCACGGCGCGCCGCAGCGGCGGAGGTGGAGCGCCCTGAGCCGCGGGCACGCGTCGAGCA 743

Qy        363 GCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCG 422
              || || ||||| ||||  |  ||     |  | ||  ||||
Db        742 CCTGGCGGTAGAAGGCGAGGGGCGTCCGCGGGTGCGTGTCC------------------- 702

Qy        423 CGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGC 482
              |  | |  ||  | ||  ||  |||| |||| |||                 | ||
Db        701 --TGCAGGACGAGCTTGGTGAGCGCGGCGTAGTGCCGGAGCTCGGACGGCGGCGGGAGGA 644

Qy        483 AGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCA 542
              || | | ||       |||| ||  || |    || || || |  || ||  || |||
Db        643 AGCAGTTGGACACGGCGAGGCTGCGGATGCGCGACCGGAGCGACTCCGCGCTGAGGCGGT 584

Qy        543 GCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGA 602
              |  |  |    |||||  | || || |  ||||| |||    | | |   |||
Db        583 GGTCGGGGAAGCAGAAGGCGTAGCAGGGGCCCTGCTCGACCAGCGCCGGCTTGACGACGG 524

Qy        603 GCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCT 662
              ||     |   |  ||  | | | |    |||| ||  |              |
Db        523 TGATGTCGAGGTCCTGGACGCCCCACGCGCCGACCGCCATGGCGATCAGGTCCTCGACGA 464

Qy        663 TGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGG 722
              | |  |   |  | |||  | |  |||   |   | | |   |||  | |  |||
Db        463 AGCCCGTGCTGCGCGCCGCGAAGAACTCGAGCCGGAGACTCCTGGCGCGGCGCGGGGCGA 404

Qy        723 TCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGT 782
              |||  |  ||  | ||||     | |  |   | |  |||  |   | | || || |
Db        403 CGGCGACGCGCGGCTCCATGAAGGCGCTGACGCAGTCGGCGAACCCGCGGATGGCGCGGC 344

Qy        783 GATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGC 842
              | |||     |  ||    ||      |||||| |  ||| || ||| |  | ||| |  ||
Db        343 GCTCGCACCGGCCGATCGCCGCGTCGAGCTTCCTGGCGTCCGCGCCGGCGGCGTCGCCGT 284

Qy        843 GGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCG 898
              ||| |||||    |  |  |  |    ||| | | |   || |  ||    || ||
Db        283 GGGCGCGGACGCAGGCGTGGTAGCGCGGCGGGAGCACGTCGCGGACGCTGAACACG 228
```

RESULT 3
US-10-156-761-6757/c
; Sequence 6757, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6757
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1431)
US-10-156-761-6757

Query Match 9.1%; Score 82.4; DB 6; Length 1431;
Best Local Similarity 45.9%; Pred. No. 5.7e-11;
Matches 410; Conservative 0; Mismatches 466; Indels 18; Gaps 3;

```
Qy          8 GCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGC 67
              ||   | ||  ||    |  || ||    |  |||| || |  ||  |||| |   | |
Db        971 GCGAGCGGGAGGGCCTTGAGCGTCGCCATCAGGCCGGTCAGATTGCCGATCAGGCGGCCC 912

Qy         68 CGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGG 127
                ||    |||||  |||  | |  | |  ||  | ||  |   |  | || || | |
Db        911 GACTTGCCGCGCGCCAGCTCGGCGATGTCCGGGTTCTTCTTCTGCGGCATGATCGACGAG 852

Qy        128 CAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCG 187
              | |  || ||    |||| |||  || |  || || |||   |  ||   |||||  |
Db        851 CCGGTGGAGAAGGCGTCGTGGAGGGTGACGAAGGAGAACTCCTTCGTGTTCCAGATGATG 792

Qy        188 TGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCC 247
                | | ||  ||  |||      |  | ||||   ||  ||  |  || |   |  |
Db        791 ACCTCCTCGGCGATCCGGGAGAGGTTCACCCCAATCATCGCGGTGATGAAGGCGAACTCG 732

Qy        248 TCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGG 307
              || ||  | |  |    ||| |||||||  |      | |   ||   || | |||  |
Db        731 GCGACGAAGTCACGGGAGGCCGTGCCGTCGATGGAGTTGGCGACCGAGCCGTGCTCGAAG 672

Qy        308 ATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCG 367
              | |  |  || | |  | |   |  | ||     |      |||   | |   |  |
Db        671 CCGAGGTCCTTCGCGACCGACTCCGGGTCGAGGCCGAGGGAGGAACCGGCGAGGG---CA 615

Qy        368 CCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCC 427
              ||  ||  |     || |||||  | ||||||||||      |   ||    ||| ||
Db        614 CCCGAGCCGTACGGCGACACGGCCGTGCGCTCGTCCCACTGGCGCAGCCGCTCGGCGTCC 555

Qy        428 GCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAAC 487
              || | |||| | |    |||| ||    ||  |    ||  |   || ||  |  |
Db        554 CGGGACAGCGCCTGGACGTGCGCGAGGACGTGGTGGGCGAAGAGCACCGGCTGGGCGTGC 495

Qy        488 TCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAG 547
              | |  ||   | |  |  | || | |  |    ||||||  | |  | | | |
Db        494 TGGAGGTGCGTGCGGCCGGGCATCGCCACGTCCGGGTGGGCCTCCGCCAGTCCGATCAGC 435

Qy        548 AACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCC 607
                | |  | ||  |  | || ||    |||| |   | | | |||||  |  |
Db        434 GCGTCCTGGAGGTCGGCGAGGAGGCCGCCGATGACGCGGGCGTGGTCGCGCAGGTACATG 375
```

```
Qy        608 CGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCG 667
              ||||| || | | |  |||      ||  |  | ||     |     | | |||||| ||
Db        374 CGGAAGAGGGTCGCGACCT------GGTCGTTCCGGGAGCGGCCGGCGCGGAGCTTGCCG 321

Qy        668 ATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACG 727
                | |  |||           |||  | | ||  ||   |  |||||     |||||   |
Db        320 CCGAGGTCGGG---------GCCGAGGCGCTCCAGCAGGCCGCGCTCCAGGGCGGTGTGG 270

Qy        728 ATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCG 787
              | ||||  |   |   |  | ||  ||  ||  |  ||  ||  | |
Db        269 ACGTCCTCGTCGGCGATGGTGCCCACGAAGGTGCCGTCCGCGACATCGGCTTCGAGCCGG 210

Qy        788 ACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGG 847
              ||||  |   |   |     |  | ||||| | || | ||||  |       | | |  |
Db        209 TCGAGGCCCGCGAGCATACGGTCCAGCTCGTCGTCCGTGAGGAGGTTCGCCTTGTGCAGC 150

Qy        848 CGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTCG 901
                | || || ||||    |  |||  |    |||||||  || |  ||||
Db        149 ACGCGCGCGTGGGCACGCGAACCTGCGATGTCGTAGGGCGCGAGCCGCCAGTCG 96
```

RESULT 4
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 9.1%; Score 82.4; DB 6; Length 9025608;
Best Local Similarity 45.9%; Pred. No. 2.1e-11;
Matches 410; Conservative 0; Mismatches 466; Indels 18; Gaps 3;

```
Qy          8 GCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGC 67
              ||   | ||  ||    |  || ||    |  |||| || |  ||  |||| |   | |
Db    8097165 GCGAGCGGGAGGGCCTTGAGCGTCGCCATCAGGCCGGTCAGATTGCCGATCAGGCGGCCC 8097106

Qy         68 CGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGG 127
                ||    |||||  |||  | |  | |  ||  | ||  |   |  | || || | |
Db    8097105 GACTTGCCGCGCGCCAGCTCGGCGATGTCCGGGTTCTTCTTCTGCGGCATGATCGACGAG 8097046

Qy        128 CAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCG 187
              | |  || ||    |||| |||  || |  || || |||   |  ||   |||||  |
Db    8097045 CCGGTGGAGAAGGCGTCGTGGAGGGTGACGAAGGAGAACTCCTTCGTGTTCCAGATGATG 8096986

Qy        188 TGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCC 247
                | | ||  ||  |||      |  | ||||   ||  ||  |  || |   |  |
Db    8096985 ACCTCCTCGGCGATCCGGGAGAGGTTCACCCCAATCATCGCGGTGATGAAGGCGAACTCG 8096926
```

```
Qy        248 TCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGG 307
Db    8096925 GCGACGAAGTCACGGGAGGCCGTGCCGTCGATGGAGTTGGCGACCGAGCCGTGCTCGAAG 8096866

Qy        308 ATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCG 367
Db    8096865 CCGAGGTCCTTCGCGACCGACTCCGGGTCGAGGCCGAGGGAGGAACCGGCGAGGG---CA 8096809

Qy        368 CCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCC 427
Db    8096808 CCCGAGCCGTACGGCGACACGGCCGTGCGCTCGTCCCACTGGCGCAGCCGCTCGGCGTCC 8096749

Qy        428 GCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAAC 487
Db    8096748 CGGGACAGCGCCTGGACGTGCGCGAGGACGTGGTGGGCGAAGAGCACCGGCTGGGCGTGC 8096689

Qy        488 TCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAG 547
Db    8096688 TGGAGGTGCGTGCGGCCGGGCATCGCCACGTCCGGGTGGGCCTCCGCCAGTCCGATCAGC 8096629

Qy        548 AACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCC 607
Db    8096628 GCGTCCTGGAGGTCGGCGAGGAGGCCGCCGATGACGCGGGCGTGGTCGCGCAGGTACATG 8096569

Qy        608 CGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCG 667
Db    8096568 CGGAAGAGGGTCGCGACCT------GGTCGTTCCGGGAGCGGCCGGCGCGGAGCTTGCCG 8096515

Qy        668 ATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACG 727
Db    8096514 CCGAGGTCGGG---------GCCGAGGCGCTCCAGCAGGCCGCGCTCCAGGGCGGTGTGG 8096464

Qy        728 ATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCG 787
Db    8096463 ACGTCCTCGTCGGCGATGGTGCCCACGAAGGTGCCGTCCGCGACATCGGCTTCGAGCCGG 8096404

Qy        788 ACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGG 847
Db    8096403 TCGAGGCCCGCGAGCATACGGTCCAGCTCGTCGTCCGTGAGGAGGTTCGCCTTGTGCAGC 8096344

Qy        848 CGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTCG 901
Db    8096343 ACGCGCGCGTGGGCACGCGAACCTGCGATGTCGTAGGGCGCGAGCCGCCAGTCG 8096290
```

RESULT 5
US-10-458-201-13/c
; Sequence 13, Application US/10458201
; Publication No. US20030207411A1
; GENERAL INFORMATION:
; APPLICANT: Jensen, Susan E
; APPLICANT: Aidoo, Kwamena A
; APPLICANT: Paradkar, Ashish S
; TITLE OF INVENTION: DNA SEQUENCE ENCODING ENZYMES OF CLAVULANIC ACID
; TITLE OF INVENTION: BIOSYNTHESIS
; FILE REFERENCE: P57425US5
; CURRENT APPLICATION NUMBER: US/10/458,201
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: US 09/385,040
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: US 08/790,462
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/567,801
; PRIOR FILING DATE: 1995-12-06
; PRIOR APPLICATION NUMBER: US 08/134,018
; PRIOR FILING DATE: 1993-10-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 11604
; TYPE: DNA
; ORGANISM: Streptomyces clavuligerus
US-10-458-201-13

Query Match 9.0%; Score 81.2; DB 6; Length 11604;
Best Local Similarity 45.1%; Pred. No. 8.9e-11;
Matches 342; Conservative 0; Mismatches 413; Indels 3; Gaps 1;

```
Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
Db       1081 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 1022

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
Db       1021 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 962

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
Db        961 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 902

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
Db        901 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 842

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
Db        841 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 782

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
Db        781 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 722

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
Db        721 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 662

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
Db        661 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 602

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
Db        601 GGAGCTCCACGGCGTACAAGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 542

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
Db        541 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 482

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
Db        481 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 422

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
Db        421 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 362

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCG 868
Db        361 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCGATGGGG 324
```

RESULT 6
US-10-939-036-35
; Sequence 35, Application US/10939036
; Publication No. US20050142094A1
; GENERAL INFORMATION:
; APPLICANT: Kumar, Manoj
; TITLE OF INVENTION: Use of Repeat Sequence Protein Polymers
; TITLE OF INVENTION: in Personal Care Compositions
; FILE REFERENCE: GC792-4-2
; CURRENT APPLICATION NUMBER: US/10/939,036
; CURRENT FILING DATE: 2004-09-10
; PRIOR APPLICATION NUMBER: US 10/800,179
; PRIOR FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/454,077
; PRIOR FILING DATE: 2003-03-12

```
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 3024
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: DNA encoding silk-elastin repeat sequence
US-10-939-036-35

  Query Match             8.9%;  Score 80.2;  DB 9;  Length 3024;
  Best Local Similarity   47.0%;  Pred. No. 1.8e-10;
  Matches  422;  Conservative    0;  Mismatches  458;  Indels   17;  Gaps    5;

Qy          5 GGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCG 64
              ||| ||   |||| |  ||| || || ||||| || |   |    | |      ||| |
Db        531 GGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCC 590

Qy         65 GGCCGCTGAGGTCGCGCGGGC--TACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCG 122
              || ||  | |   |  | |||    | ||||| |||| || | |||   |||   |   |
Db        591 GGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGG 650

Qy        123 CCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGA 182
              | ||   |      |          | ||  |   ||| |  || |  | ||  ||  |   |
Db        651 CGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCA 710

Qy        183 TGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCA 242
              || |  | || ||   |||  | ||  ||||   |    | ||  | |||| |  |  |
Db        711 TGGCCCGGCGGGCCCGAAAGGCGCGCATGGCCCGGCGGGCCCGAAAGGCGCGCATGGCCC 770

Qy        243 CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGAT 302
              |    | |    | || ||   |||  || | |||  |    |   |   |  |||  |||
Db        771 GGCGGGCCCGAAAGGCGCGCATGGCCCGGCGGGCCCGAAAGGCGCGCAGGGCCCGGCGGG 830

Qy        303 CGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCG 362
              | |||    ||||| | |  | ||   |    || ||| |||    | | ||||||   |
Db        831 CCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCG-CGCAGGGCCCGGCGGGCCCGG 889

Qy        363 GCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCG 422
              ||    || || |  ||  | ||| |      |  ||  |  | | || |  |||
Db        890 GCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCG 949

Qy        423 CGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGC 482
              |||  |  | |   |  |||||||  | ||  |  ||||    |    ||  ||
Db        950 CGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCC---GGGCGGCGCGC 1006

Qy        483 AGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCA 542
              ||  | |||||    || |||    |  ||       | ||| || |    ||  | ||
Db       1007 AGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCC 1066

Qy        543 GCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAG------AGATAGC 596
              | | |   ||    |||| ||  | | ||  | ||    ||  |              |
Db       1067 CGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGG 1126

Qy        597 GCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGA 656
              || || ||  | | | | | | | | | | | | |  |  |  |  |  |  |  |  |
Db       1127 GCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGG 1186

Qy        657 GCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGG 716
              || ||  |  | | ||  |||| |  |    ||| || |  | |  | || | || | |
Db       1187 GCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCG 1246

Qy        717 TCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGG 776
                || | ||  |  |    ||   | |||  | |   ||   || |||||    |||
Db       1247 CGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGG 1306

Qy        777 TGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGA 836
              |||     | |  |  | |||| ||  |  | |   |||| ||   ||    ||
Db       1307 GCCCG-----GCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGG 1361

Qy        837 TGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCC 893
              |||||  |||   ||||   |||  | || |||   ||| ||  |||||
Db       1362 CCCGGCGGGCCCGGGCGGCGCGCAGGGCCCGGCGGGCCCGGGCGGCGCGCAGGGCCC 1418
```

```
RESULT 7
US-10-458-201-15/c
; Sequence 15, Application US/10458201
; Publication No. US20030207411A1
; GENERAL INFORMATION:
;  APPLICANT: Jensen, Susan E
;  APPLICANT:  Aidoo, Kwamena A
;  APPLICANT:  Paradkar, Ashish S
;  TITLE OF INVENTION: DNA SEQUENCE ENCODING ENZYMES OF CLAVULANIC ACID
;  TITLE OF INVENTION:  BIOSYNTHESIS
;  FILE REFERENCE: P57425US5
;  CURRENT APPLICATION NUMBER: US/10/458,201
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: US 09/385,040
;  PRIOR FILING DATE: 1999-08-30
;  PRIOR APPLICATION NUMBER: US 08/790,462
;  PRIOR FILING DATE: 1997-01-29
;  PRIOR APPLICATION NUMBER: US 08/567,801
;  PRIOR FILING DATE: 1995-12-06
;  PRIOR APPLICATION NUMBER: US 08/134,018
;  PRIOR FILING DATE: 1993-10-08
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
;   LENGTH: 1722
;   TYPE: DNA
;   ORGANISM: Streptomyces clavuligerus
US-10-458-201-15

  Query Match             8.9%;  Score 80;  DB 6;  Length 1722;
  Best Local Similarity   45.1%;  Pred. No. 2.2e-10;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173
              || ||    |   || || || ||    | ||| |   |  |    |  || | ||
Db        898 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 839

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233
              |   |||| ||  |   |  | ||| ||| |    |||| |     ||| ||| ||
Db        838 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 779

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290
              |||  ||| | |   |  | ||| |    ||   || |   |    |||| ||  | |
Db        778 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 719

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350
              |    ||||          |||   | || || |  | ||| |||  || ||     ||
Db        718 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 659

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||   | |||
Db        658 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 599

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |     |
Db        598 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 539

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db        538 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 479

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||          || |  ||  |    || || |  |  | |||||||  | |||
Db        478 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 419

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |    ||||  | |  | ||| ||   |  |   | ||   | | ||| |  || ||
Db        418 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 359

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
                ||      |  |  |||  ||   | | ||    |  |||  |||  |       |
```

```
Db        358 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 299

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770

Db        298 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 239

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830

Db        238 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 179

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862

Db        178 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 147


RESULT 8
US-10-458-201-1/c
; Sequence 1, Application US/10458201
; Publication No. US20030207411A1
; GENERAL INFORMATION:
;  APPLICANT: Jensen, Susan E
;  APPLICANT:  Aidoo, Kwamena A
;  APPLICANT:  Paradkar, Ashish S
;  TITLE OF INVENTION: DNA SEQUENCE ENCODING ENZYMES OF CLAVULANIC ACID
;  TITLE OF INVENTION:  BIOSYNTHESIS
;  FILE REFERENCE: P57425US5
;  CURRENT APPLICATION NUMBER: US/10/458,201
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: US 09/385,040
;  PRIOR FILING DATE: 1999-08-30
;  PRIOR APPLICATION NUMBER: US 08/790,462
;  PRIOR FILING DATE: 1997-01-29
;  PRIOR APPLICATION NUMBER: US 08/567,801
;  PRIOR FILING DATE: 1995-12-06
;  PRIOR APPLICATION NUMBER: US 08/134,018
;  PRIOR FILING DATE: 1993-10-08
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 15120
;   TYPE: DNA
;   ORGANISM: Streptomyces clavuligerus
US-10-458-201-1

  Query Match             8.9%;  Score 80;  DB 6;  Length 15120;
  Best Local Similarity   45.1%;  Pred. No. 1.7e-10;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173

Db       3154 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 3095

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233

Db       3094 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 3035

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290

Db       3034 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 2975

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350

Db       2974 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 2915

Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410

Db       2914 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 2855

Qy        411 CCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470

Db       2854 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 2795

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530

Db       2794 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 2735

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590

Db       2734 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 2675

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650

Db       2674 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 2615

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710

Db       2614 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 2555

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770

Db       2554 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 2495

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830

Db       2494 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 2435

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862

Db       2434 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 2403


RESULT 9
US-10-494-495-1/c
; Sequence 1, Application US/10494495
; Publication No. US20050079498A1
; GENERAL INFORMATION:
;  APPLICANT: SmithKline Beecham plc
;  APPLICANT:  The Governors of the University of Alberta
;  TITLE OF INVENTION: Polynucleotides and Polypeptides involved in Clavulanic
;  TITLE OF INVENTION:  Acid Biosynthesis and Use Thereof
;  FILE REFERENCE: P32913
;  CURRENT APPLICATION NUMBER: US/10/494,495
;  CURRENT FILING DATE:  2004-05-04
;  PRIOR APPLICATION NUMBER: PCT/GB02/04989
;  PRIOR FILING DATE: 2002-11-06
;  PRIOR APPLICATION NUMBER: GB 0126756.6
;  PRIOR FILING DATE: 2001-11-07
;  PRIOR APPLICATION NUMBER: GB 0128776.2
;    2001-11-30
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 29870
;   TYPE: DNA
;   ORGANISM: Streptomyces clavuligerus
US-10-494-495-1

  Query Match             8.9%;  Score 80;  DB 9;  Length 29870;
  Best Local Similarity   45.1%;  Pred. No. 1.6e-10;
  Matches  339;  Conservative    0;  Mismatches  410;  Indels    3;  Gaps    1;

Qy        114 CGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCA 173

Db       5423 CGGCGTAGTCGTAGCCGACGGTGAGGACGAGGTCCACCGGGGCGAACATGGTCTGGAGCG 5364

Qy        174 CGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGC 233

Db       5363 CCGGGAAGTTGAGGATGCCGTCCATGTAGCCGGTGACGGCGCCGTAGTTCAGCTCGTGGC 5304

Qy        234 GGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCC---GTCCCTCAGCCGGACCG 290

Db       5303 CGACCGGCAGGACACCCTTGGCGATGTAGGTCGTGATGACCGGGATGTTCAGGCGCTCGG 5244

Qy        291 ACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACA 350

Db       5243 CCAGGGCGCGGATCGCCGGGACGGCGCCCGAGCGGATCGCGGCCGCTCCGACGACGAGCA 5184
```

```
Qy        351 CCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCG 410
              ||| | ||   | |||| ||   ||||||  |  |  | |  ||  ||  | |||
Db       5183 CCGGGTGCTTGGCCTCGGCGAGCAGGGCGGCGGCCTGGTCGGCGGCCTTCTGCCAGCCGT 5124

Qy        411 CCTCGCGGCGCGCGCCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGA 470
              |  ||  |    || |||    || || | || ||| ||   |||   ||| |      |
Db       5123 CGGCGACGACGCCGACCGGTTTCGCCGGGGTGTTCGCCGGCGGGTTGGGGACGGTGGTGT 5064

Qy        471 AGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCT 530
              || || |  | || |  ||       || | || || |  |  |||   | ||| | ||
Db       5063 CGATGCCCTCGGAGGAGCCGAGCAGGTCCACCGGGAGGGAGATGAAGGAGGGCCCGACCG 5004

Qy        531 CGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              |||        || |  ||  |   || ||  |  |  |  |||||||  | |||
Db       5003 GCTCGGTCATGGCCGCGTTCACGGCGGAGTCGACGAGGTCGGTGATCTCGTGGGGCCGCT 4944

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |      ||||   | | | ||| ||   |  |   | |||   | | ||| |  || ||
Db       4943 GGAGCTCCACGGCGTACTTGGACATCGGGGCGACGATCGCCACCGAGTCCAGGCACTGGT 4884

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              ||     |  |  |||  ||   | | ||   |  |||  |||  |      |
Db       4883 GGGTGTCGTTCGGGAAGATGTCGTGCGACTCCGACTGCGCGGCGAGCGCGATGACCGGCG 4824

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              ||  |  |       || |  |      |||    |||    |    |   |  | ||
Db       4823 AGCGGTCCAGGACGGACGTGGCGATACCGGTGGAGAGGTTGGTCATACCGGGGCCCAGGG 4764

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              ||     | | | |  ||  |   |||||||   || ||  || |
Db       4763 TGGCCCAGCACGCCTGGGGGCGACCGGTGATCCGCGCGAGGACATCAGCGGCGACACCCG 4704

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCG 862
              ||| ||  ||| || ||   ||  ||  | ||
Db       4703 CGGTGAACTCGTGGCGGGTCAGAACGAAGTCG 4672


RESULT 10
US-10-282-122A-14103/c
; Sequence 14103, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14103
;   LENGTH: 5223
;   TYPE: DNA
;   ORGANISM: Burkholderia mallei
US-10-282-122A-14103

  Query Match             8.6%;  Score 77.8;  DB 7;  Length 5223;
  Best Local Similarity   47.8%;  Pred. No. 6.7e-10;
  Matches  359;  Conservative    0;  Mismatches  377;  Indels   15;  Gaps    4;

Qy        133 GGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCGTGCCG 192
              | ||| |   | ||||     | | |  || ||    || ||| ||  | |||| || |
Db       1170 GTCGACCACCTGATCGAACGGCACGTCCTGGTGCGACTGCGCGTCGATCAGGCCGCGCTG 1111

Qy        193 CGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCCTCGCC 252
              ||   |  | ||||||| |   | |      || || ||||| || || || | |
Db       1110 CGTCGCGTCGACGAGCGCGTCGAACGC-----GAGCGTCGCGGGCGTGCTCGCCGCGAGC 1056

Qy        253 GTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGGATCAC 312
              ||| ||   |  |       | |  |||||| ||| | | |  |||| ||||  || |
Db       1055 GTCAGCGTATT-GACGAAAAAGCCGATCAGCCCGACGGTCTCGCCGCGCTCGCGATTCGC 997

Qy        313 GCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCGCCGTA 372
              | ||  ||||  |  |  | | |  ||||  | | |    |   || ||    ||
Db        996 CGACGGCACGCCGATCCGGATCCCGGTCTCGCCGCTCGCGCGCGCGAGCAGCGTCGCGAG 937

Qy        373 GAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCCGCGGC 432
              ||||   | ||||  |  ||    |||    |||||      |||| | |  |
Db        936 CGCGGCGAGCAGCACGGCGAACGGCGTCGCCCGGCGCGCCTGCGCCAGCGCACGCATGCG 877

Qy        433 CGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGC 492
              | |||||      |  |  ||  |      | |  | || |  || |  ||  ||
Db        876 CCGCGCGAGTTCGGCGGATATGTCGAACGCATGGCGCGCGCCGCTCGCGCCGCGCTGCGC 817

Qy        493 GTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACAT 552
              |        || |  |  |  | | |  ||  | |  |    ||  ||||
Db        816 GGGCCGCGTCGCGGCGCCCGGCAGCAGCAGCGCGTCACGGCTCGCGTCGAGCGTCGCGCG 757

Qy        553 CGGATAGGCGCGGGAGCGGTCGATCT------CGACGAGCCAGAGATAGCGCACGAGCTC 606
              |   || ||| |   ||| |||  ||      || ||||||||  |     ||  |||| |
Db        756 CCAGTACGCGAGCTGGCGCTCGCCCTCGCCCGCGTCGAGCCAGCGGCGCTGCCAGAGCGC 697

Qy        607 CCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTC 666
              |  |     ||  | | | | |  |      |  ||      |  |
Db        696 GTAGTCCGCGTACTGCACCGGCAGCGGCGCGAGCGACACCGCGTCGCCCGTCGCGTACGA 637

Qy        667 GATCACGAAGGGCGCCCGGTAAGCGAAGAG---ATCAGGCGTCCTGCGCTGGGTCGCGGT 723
              |||  | || | | |  |||| |    || | || || ||  |  | | || |
Db        636 CCGGTAGAACGACGACAGCTCCGCGAGCATCACGTCGGCCGACCAGCCGTCCGACACGAT 577

Qy        724 CACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTG 783
              |      |||  ||| |   |||| |      || ||||||  |    | |    || |
Db        576 GTGGTGCGTCGTCAGCACGAACCAGTGCGTGGCCGCATCGAGCCGGATCAGGTGCGCGCG 517

Qy        784 ATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCG 843
              |||    || | ||||  ||| | ||||||||||| || |   ||| |||| ||
Db        516 CACGAGCGGGCCGTGCGCGAGATCGAACGGCTCGGCCTCGTCGCGCTCGGCGATGGCGGC 457

Qy        844 GGGGCGGAGCTCGCGGGCGGCGGCGAGCGCG 874
              | |||  ||||||| |   |||||  ||||
Db        456 CGCGCGCTGCTCGCGCGACCCGGCGCTCGCG 426


RESULT 11
```

```
US-10-084-846A-45/c
; Sequence 45, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
;  APPLICANT: WEITNAUER, GABRIELE
;  APPLICANT:  MUHLENWEG, AGNES
;  APPLICANT:  TREFZER, AXEL
;  APPLICANT:  BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 45
;   LENGTH: 1431
;   TYPE: DNA
;   ORGANISM: Streptomyces viridochromogenes
;   FEATURE:
;   OTHER INFORMATION: aviJ dna: partial sequence of coding strand 1; nucleotide 1
;   OTHER INFORMATION: corresponds to nucleotide 15,012 of coding strand 1.
US-10-084-846A-45

  Query Match             8.6%;  Score 77.4;  DB 6;  Length 1431;
  Best Local Similarity   45.0%;  Pred. No. 9.8e-10;
  Matches  380;  Conservative    0;  Mismatches  456;  Indels    9;  Gaps    2;

Qy         60 GCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGA 119
              | |||  |  | | |   |||||| || ||||  || |  |     | |  ||||| |||
Db       1006 GGGCGTCCGCCAGGGCCGGCGCGGTCTTCGGGCAGATGGCGCCGAACAGGAAGGCGCCGA 947

Qy        120 GCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGA 179
                 |   || | | ||  | |||  |  | |    |     | ||  |||    |  ||
Db        946 AGATGAGATGGATGCCTATGAGCTCGGTGGCGACCGCGGACAGCAGGACGCCGGGGATGA 887

Qy        180 AGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGA 239
               |    | ||    ||  |  | | |||  |  ||||  ||||     ||||||    |
Db        886 TCAGTGTGAGCACGCCGCCGTCGGGCAGCCGGCGCTCCGGCCGCTCCAGGACGCGTCGCA 827

Qy        240 GCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGC 299
                |||        |  ||||||| || || || || |     |       | |    | ||
Db        826 AGGCGG------GGCGGACGAGCACCACCATGACGCCGAAGAAGACCGCGGTGAGCGCGA 773

Qy        300 GATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCG 359
              |   ||| | ||||| | | ||     |   ||  |  |  | ||        ||||| ||
Db        772 GGGTGCGCAGCACGCCCAGGGGCGAACCGCCGCGCAACAGGGCGACGAGACCCGCGAGCA 713

Qy        360 GCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGC 419
              || ||  || |  |||  ||  |   ||||||  | |    | |  | ||  || | |||
Db        712 GCAGCCAGCAGGCGAGATCGGCGGTGGCGGCCCCCACGATGGCGTAGGTGCCCGCCCGGG 653

Qy        420 GCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCC 479
              |  |  |     |  |  |   | ||        |||  |  |     ||    | || |
Db        652 ACTGGAGCATCCCGCGTTCCATCAGCAGCCGCGCCAGCACGGGGAAGGCGGTCACGCTCA 593

Qy        480 GGCAGAACTCGGCGTATT---CGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGA 536
              |  ||   |   | |     || ||  | || | |||    || |  ||| | | ||
Db        592 TGGCGATGCCCAGGAAGAGGGTGAACGGCAGGAAGCCGACGCCGTCCGGGGCGTAGCCGG 533

Qy        537 CGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGC 596
              |  |||||  | |   | ||     || |  |  |     | ||  ||  | ||     |
Db        532 TGTACAGCGCGTATCCCAGAGCACAGCCGAGGGCGAACGACACGCAGACGCCGACGATTC 473

Qy        597 GCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGA 656
              ||||    |||  |     |  || |  | || |    |||||  |  ||  ||   |  ||
Db        472 CCACGGCGGCCCCGGTGCGGCCCGCGCCTTCATGGTGGGTGTCACCGAACTCCATGCCGA 413

Qy        657 GCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGG 716
                  |||      || ||      || |    |  |||| ||||||  |     | |  |||

Db        412 TCAGGAACATGAACAGCGCCAGCCCGATCTGGGCGACGAGATCGAGGTAGGAGTGGAGGG 353

Qy        717 TCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGG 776
                |||| | || | |    || |              |  || |   | ||   |   |
Db        352 CGGCGGGGAAGAGGGCGTCGTGGACCCCGGGAGCGACGAGGCCGAGCACCGAAGGGCCGA 293

Qy        777 TGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGA 836
                ||    ||  || |     | | | | || |  ||| | ||  |    |||||   |
Db        292 GAACGATCCCGGAGATCATCTCGCCCACCACGCGTGGCTGGCCCAAGTGCCGGCCGACCA 233

Qy        837 TGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACG 896
                | | |  |    || ||  |||  |  | |  ||| ||        ||     |||||
Db        232 GCGCTCCGCAGGCCCGCGCGACGGCCACCACCACGGCGAGGGCGAGCAGCAACCGCCACG 173

Qy        897 CGTCG 901
                |||
Db        172 CCTCG 168


RESULT 12
US-10-084-846A-1/c
; Sequence 1, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
;  APPLICANT: WEITNAUER, GABRIELE
;  APPLICANT:  MUHLENWEG, AGNES
;  APPLICANT:  TREFZER, AXEL
;  APPLICANT:  BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
;   LENGTH: 59816
;   TYPE: DNA
;   ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

  Query Match             8.6%;  Score 77.4;  DB 6;  Length 59816;
  Best Local Similarity   45.0%;  Pred. No. 6.5e-10;
  Matches  380;  Conservative    0;  Mismatches  456;  Indels    9;  Gaps    2;

Qy         60 GCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGA 119
              | |||  |  | | |   |||||| || ||||  || |  |     | |  ||||| |||
Db     16017 GGGCGTCCGCCAGGGCCGGCGCGGTCTTCGGGCAGATGGCGCCGAACAGGAAGGCGCCGA 15958

Qy        120 GCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGA 179
                 |   || | | ||  | |||  |  | |    |     | ||  |||    |  ||
Db     15957 AGATGAGATGGATGCCTATGAGCTCGGTGGCGACCGCGGACAGCAGGACGCCGGGGATGA 15898

Qy        180 AGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGA 239
               |    | ||    ||  |  | | |||  |  ||||  ||||     ||||||    |
Db     15897 TCAGTGTGAGCACGCCGCCGTCGGGCAGCCGGCGCTCCGGCCGCTCCAGGACGCGTCGCA 15838

Qy        240 GCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGC 299
                |||        |  ||||||| || || || || |     |       | |    | ||
Db     15837 AGGCGG------GGCGGACGAGCACCACCATGACGCCGAAGAAGACCGCGGTGAGCGCGA 15784

Qy        300 GATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCG 359
              |   ||| | ||||| | | ||     |   ||  |  |  | ||        ||||| ||
Db     15783 GGGTGCGCAGCACGCCCAGGGGCGAACCGCCGCGCAACAGGGCGACGAGACCCGCGAGCA 15724

Qy        360 GCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGC 419
              || ||  || |  |||  ||  |   ||||||  | |    | |  | ||  || | |||
Db     15723 GCAGCCAGCAGGCGAGATCGGCGGTGGCGGCCCCCACGATGGCGTAGGTGCCCGCCCGGG 15664

Qy        420 GCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCC 479
```

```
Db      15663 ACTGGAGCATCCCGCGTTCCATCAGCAGCCGCGCCAGCACGGGGAAGGCGGTCACGCTCA 15604

Qy        480 GGCAGAACTCGGCGTATT---CGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGA 536

Db      15603 TGGCGATGCCCAGGAAGAGGGTGAACGGCAGGAAGCCGACGCCGTCCGGGGCGTAGCCGG 15544

Qy        537 CGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGC 596

Db      15543 TGTACAGCGCGTATCCCAGAGCACAGCCGAGGGCGAACGACACGCAGACGCCGACGATTC 15484

Qy        597 GCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGA 656

Db      15483 CCACGGCGGCCCCGGTGCGGCCCGCGCCTTCATGGTGGGTGTCACCGAACTCCATGCCGA 15424

Qy        657 GCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGG 716

Db      15423 TCAGGAACATGAACAGCGCCAGCCCGATCTGGGCGACGAGATCGAGGTAGGAGTGGAGGG 15364

Qy        717 TCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGG 776

Db      15363 CGGCGGGGAAGAGGGCGTCGTGGACCCCGGGAGCGACGAGGCCGAGCACCGAAGGGCCGA 15304

Qy        777 TGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGA 836

Db      15303 GAACGATCCCGGAGATCATCTCGCCCACCACGCGTGGCTGGCCCAAGTGCCGGCCGACCA 15244

Qy        837 TGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACG 896

Db      15243 GCGCTCCGCAGGCCCGCGCGACGGCCACCACCACGGCGAGGGCGAGCAGCAACCGCCACG 15184

Qy        897 CGTCG 901

Db      15183 CCTCG 15179


RESULT 13
US-10-084-846A-2
; Sequence 2, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
;  APPLICANT: WEITNAUER, GABRIELE
;  APPLICANT:  MUHLENWEG, AGNES
;  APPLICANT:  TREFZER, AXEL
;  APPLICANT:  BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
;   LENGTH: 59816
;   TYPE: DNA
;   ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

  Query Match             8.6%;  Score 77.4;  DB 6;  Length 59816;
  Best Local Similarity   45.0%;  Pred. No. 6.5e-10;
  Matches  380;  Conservative    0;  Mismatches  456;  Indels    9;  Gaps    2;

Qy         60 GCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGA 119

Db      43800 GGGCGTCCGCCAGGGCCGGCGCGGTCTTCGGGCAGATGGCGCCGAACAGGAAGGCGCCGA 43859

Qy        120 GCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGA 179

Db      43860 AGATGAGATGGATGCCTATGAGCTCGGTGGCGACCGCGGACAGCAGGACGCCGGGGATGA 43919

Qy        180 AGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGA 239

Db      43920 TCAGTGTGAGCACGCCGCCGTCGGGCAGCCGGCGCTCCGGCCGCTCCAGGACGCGTCGCA 43979

Qy        240 GCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGC 299

Db      43980 AGGCGG------GGCGGACGAGCACCACCATGACGCCGAAGAAGACCGCGGTGAGCGCGA 44033

Qy        300 GATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCG 359

Db      44034 GGGTGCGCAGCACGCCCAGGGGCGAACCGCCGCGCAACAGGGCGACGAGACCCGCGAGCA 44093

Qy        360 GCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGC 419

Db      44094 GCAGCCAGCAGGCGAGATCGGCGGTGGCGGCCCCCACGATGGCGTAGGTGCCCGCCCGGG 44153

Qy        420 GCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCC 479

Db      44154 ACTGGAGCATCCCGCGTTCCATCAGCAGCCGCGCCAGCACGGGGAAGGCGGTCACGCTCA 44213

Qy        480 GGCAGAACTCGGCGTATT---CGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGA 536

Db      44214 TGGCGATGCCCAGGAAGAGGGTGAACGGCAGGAAGCCGACGCCGTCCGGGGCGTAGCCGG 44273

Qy        537 CGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGC 596

Db      44274 TGTACAGCGCGTATCCCAGAGCACAGCCGAGGGCGAACGACACGCAGACGCCGACGATTC 44333

Qy        597 GCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGA 656

Db      44334 CCACGGCGGCCCCGGTGCGGCCCGCGCCTTCATGGTGGGTGTCACCGAACTCCATGCCGA 44393

Qy        657 GCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGG 716

Db      44394 TCAGGAACATGAACAGCGCCAGCCCGATCTGGGCGACGAGATCGAGGTAGGAGTGGAGGG 44453

Qy        717 TCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGG 776

Db      44454 CGGCGGGGAAGAGGGCGTCGTGGACCCCGGGAGCGACGAGGCCGAGCACCGAAGGGCCGA 44513

Qy        777 TGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGA 836

Db      44514 GAACGATCCCGGAGATCATCTCGCCCACCACGCGTGGCTGGCCCAAGTGCCGGCCGACCA 44573

Qy        837 TGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACG 896

Db      44574 GCGCTCCGCAGGCCCGCGCGACGGCCACCACCACGGCGAGGGCGAGCAGCAACCGCCACG 44633

Qy        897 CGTCG 901

Db      44634 CCTCG 44638


RESULT 14
US-10-156-761-1868/c
; Sequence 1868, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
```

```
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1868
;   LENGTH: 1491
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1491)
US-10-156-761-1868

  Query Match             8.5%;  Score 77;  DB 6;  Length 1491;
  Best Local Similarity   45.4%;  Pred. No. 1.2e-09;
  Matches  406;  Conservative    0;  Mismatches  475;  Indels   14;  Gaps    3;

Qy          3 CGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCG 62
              ||   || | || |||   | |  || |||| ||||       |  ||  |        ||
Db       1049 CGCTCGCGTACCAGCTCCGGATGGACGCCGAGACCGACGGAGCCGAGGAAGCAGCGCTCG 990

Qy         63 CGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCG 122
                | || | |||||   |||| | |||       |||   |||||  |  | ||  | |||
Db        989 ATGTCCCCGGAGGTGAAGCGGCCGACGTCCACGCGCACGGCGTCGCCCTGCCGCAGGGCG 930

Qy        123 CCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGA 182
              |  | |||    ||  |  | | |||| | ||||| || |       | |  ||    ||
Db        929 CGGGTCAGGTCACGTAC--GCCCTCGATGCCGAGGTCGTGCGCGAGGTGGTCGGGTGAGC 872

Qy        183 TGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCA 242
              |||  || || ||  ||   ||||  ||      |      ||  ||| ||  | ||  ||
Db        871 CGCCCGGCAGCACGGCGAGCGGGAGGCCGTGACGCAGGGCCACCTCGGCGGCGGCGTTCA 812

Qy        243 CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGAT 302
              | |   || ||||  ||  | || |     | |       |  | | | |  |||    |
Db        811 CCGTGCCGTCGTCCCCGCACACCCCGAGCACCTGGGCGCGGGCGGCCGGCCTTCTCCAACT 752

Qy        303 CGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCG 362
              ||     |||      |||  |||||     || | || | |     |  |||  |
Db        751 CGGCCCGCACCTCGTCCGGTTCGCACTGCACCGTCTCGGCACGGGGGAGGGCGTCCCCGA 692

Qy        363 GCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTC------GCCTCGC 416
              |  |||  |  || |   |  ||| ||  |  | ||| ||      |        || ||||
Db        691 GGGCGCGGACGCGGTCCGGGGCCGCCGCGCCCTGGTCGGCCACGACCACCAGACCCGCGC 632

Qy        417 GGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGC 476
              |   |     ||||| ||    |  || |||   |    || |   | ||| |  |  |
Db        631 CGTCGGGCAGCGCGGGCGCGTCCACGCGCGGGCGGCCCGGCGGCGGGAGCCGGTCGCGGG 572

Qy        477 GCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGA 536
              | |  || | ||  |    ||| || |  |     ||  | ||  | | ||| |   ||
Db        571 TCGGCACGAGCCCGCGGACGGCGAACGCGGCGCCCACGCCCAGGGCCGCGCCCGCCAGGA 512

Qy        537 CGCGCAGCGAGAA------CATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGA 590
              ||     || |||          ||  ||| |||| ||| || |||   || ||
Db        511 CGTCGCTCGGGAAGTGGACGCCCGTGTAGACGCGCGAGAGGGCGACGGCGGCGGCGAGCG 452

Qy        591 GATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGT 650
              |   | || |     || ||   | | |||     | | |  || || | ||    |  |
Db        451 GTGCGAGCGCCGCGCCCAGGCCGCGTGACTCCAGGGCGACGCCCGTCGCGAAGGCGGCGG 392

Qy        651 TCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGC 710
              |  |     | | |   | ||| | ||    |  |     | | |   |     | | |
Db        391 CCGAGGCGGAGTGGCCCGACGGGAACGACGTGGTGATCGGCTGGCGCTTGAGGCGGCGGA 332

Qy        711 GCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCT 770
              || |   | |   |||   | |   ||  | | | |  |     |      |||| |  |
Db        331 CGTGCGGGACCGGGTCGAGAACGGGGCGCGGGCGGCGGACCGAGCGCTTGCCGAGGGTGT 272

Qy        771 GAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGC 830
              |  ||| ||    | | |      |||||  |  |     |||||  |  |    |
Db        271 TGATCGTGGCGGAGGCCAGGGTGAGCGAGGCGAGGCCCCGGGCGGCGGCCCGGCGGGCAC 212

Qy        831 CGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGG 885
              |||| || ||  ||    |||  ||  ||| ||||||| |    ||| | || |
Db        211 GGGGGCTGCCGGTGGCCGCGAGGGCGGTGGCCGCGGCGAACCACAGGACCCCGTG 157
```

```
RESULT 15
US-09-861-289-32/c
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             8.3%;  Score 75.2;  DB 3;  Length 11220;
  Best Local Similarity   45.2%;  Pred. No. 2.7e-09;
  Matches  409;  Conservative    0;  Mismatches  478;  Indels   17;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              || ||| ||| || |   ||| |  | |   | |  ||| |       |  || | ||||
Db      10120 GGAGGGTGCCGTCGGAGAGGACGCCTGCCGTGTGGACGACCGCGGTGAGCGGGTGTTCGG 10061

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              || ||  ||     |   ||  ||  |  |||| |||||    |        |    ||
Db      10060 CGGGGATCGAGTCGAGTACGGCGGTGA--GGGCTTCGCGGTCGGCGACGTCGCACGCGGC 10003

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              | ||  |  |  |||   | || | || || ||| |  || |     ||    |   |
Db      10002 CACCGAGACGTCGGCTCCCAGGGCCTCCAGCTCGTGCACGAGCTCGCCGGCGCCCGGGGC 9943

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              |      | ||||  | | | | | |||  |       | | |   ||| |    ||| |
Db       9942 GTCCGTGCCCCGCCGGCTCACGAGCAGCAGGCGTCGTACGCCCCACTCGCCCACCACGTG 9883

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              | |||| ||  | |  | ||| |  |  ||||| |  ||||| |||||    |||| |
Db       9882 CCGGGCCACGATGCCCCCCAGCGCACCGGTGCCGCCGGTCAGCAGGACCGTACCCTCCGG 9823

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              ||| ||  |    ||  || | |  ||     ||| || ||                | |
Db       9822 GTCGAGGCCCGACGGCATCGTGAGGACGACCTTGCCCGTGTGGC------------GGGC 9775

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              | |   | |||   ||  |  ||    ||||  | || ||||||| ||||   || | |
Db       9774 CTGGCTGACGTGCCGGAAGGCGTCGCGGGCCCGGCGCACGTCCCAGGTCGTGACGGGCAG 9715

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
                    |    || | |  | ||| |  |  |  |  | |||     |  | |||
Db       9714 GTGCCGGAGCACCCCGTCCTCGAAGAGGGCGATGACCTCGGCGAGCATCTCGCCGATCCG 9655

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
                  | |||| |   | ||   ||||  |  | |   |  | ||  | |||| ||
Db       9654 CTCCGGCCCGGCCTCGCCCAGGTCGAAGGCGCGGTAGCCGACACCGGGGTGGTCGGCGGC 9595

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              | |      |          | ||  ||      | ||||| |||| || |        || |
Db       9594 GACCCGCTCCGCGTCGCGGACGTCGGTCTTCCCCATCTCCACGAACCGGCCGCCCGGCCC 9535

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660
              ||||   || | |   || ||  |   ||| |||| |  |  | ||      | |  ||
Db       9534 GAGCAGGCGCAGCGAGGCGTCGACGAACTCGCGGGCGAGCGAGTTCAGTACGACGTCCAT 9475
```

```
Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAA---GAGATCAGGCGTCCTGCGCTGGGT 717
              |      || | |  || ||| | ||| |   ||  ||  | || | |     ||
Db       9474 GCCCGCCCCGCCGGAAGCGGCACGGAACGCGGACTCGAAGTCCAGGGTGCGGGAGGAGGC 9415

Qy        718 CGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGT 777
              |  | ||  ||||      |  |  |      | |      |       |       ||
Db       9414 GATGTGCGCGTCGTCCAGGCCGAGCGCGCGCAGGGCGTCCCACTTCCCGTGACTCGCCGT 9355

Qy        778 GCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGAT 837
              ||||||  |  | |   | | |    | |||||||| ||||  |    |    ||
Db       9354 GCCGTGGACCTCCACGCCCCAGTGCCGGGCGAGCTGCACGGCGGCCATGCCCACGCCACC 9295

Qy        838 GTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGC 897
              | ||  || | ||| |  | ||    || || ||  |  |  |||  |  ||  |   ||
Db       9294 GGCGGCGGAGTGGACCAGGAGGCGCTCGCCGGGCTTGACGTCCGCCAGGTCGCGCAGGGC 9235

Qy        898 GTCG 901
              || |
Db       9234 GTAG 9231
```

Search completed: April 27, 2006, 14:40:03
Job time : 908.317 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 565.78 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_76000_76900
Perfect score: 901
Sequence: 1 ggcgggggccttccggctgg.....cgggcgcggcccacgcgtcg 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	78	8.7	9222	7	US-10-496-351-2	Sequence 2, Appli
c	2	78	8.7	82746	7	US-10-496-351-56	Sequence 56, Appl
c	3	77.8	8.6	9222	7	US-10-496-351-2	Sequence 2, Appli
	4	77.8	8.6	82746	7	US-10-496-351-56	Sequence 56, Appl
c	5	75.4	8.4	4770	7	US-10-496-351-3	Sequence 3, Appli
c	6	72	8.0	113193	7	US-10-478-943D-1	Sequence 1, Appli
c	7	71.8	8.0	14055	7	US-10-496-351-1	Sequence 1, Appli
c	8	71.6	7.9	1773	11	US-11-096-568A-22025	Sequence 22025, A
c	9	69.4	7.7	2742	7	US-10-496-351-22	Sequence 22, Appl
c	10	69	7.7	88421	14	US-11-205-109-1	Sequence 1, Appli
c	11	68.4	7.6	1339	11	US-11-096-568A-27264	Sequence 27264, A
c	12	68	7.5	9975	7	US-10-496-351-4	Sequence 4, Appli
c	13	67.6	7.5	3354	7	US-10-496-351-12	Sequence 12, Appl
c	14	67.4	7.5	11905	7	US-10-496-351-57	Sequence 57, Appl
c	15	65.8	7.3	3305	14	US-11-017-550-68	Sequence 68, Appl
c	16	65.8	7.3	12391	14	US-11-017-550-67	Sequence 67, Appl
	17	64	7.1	11070	14	US-11-075-185-34	Sequence 34, Appl
	18	64	7.1	78869	14	US-11-075-185-1	Sequence 1, Appli
c	19	63.8	7.1	984	11	US-11-228-659-45	Sequence 45, Appl
c	20	63.8	7.1	4470	11	US-11-228-659-44	Sequence 44, Appl
	21	63.4	7.0	9975	7	US-10-496-351-4	Sequence 4, Appli
	22	63.4	7.0	14055	7	US-10-496-351-1	Sequence 1, Appli
	23	63.4	7.0	88421	14	US-11-205-109-1	Sequence 1, Appli
	24	62	6.9	10968	14	US-11-075-185-35	Sequence 35, Appl
c	25	61.8	6.9	1676	11	US-11-096-568A-20689	Sequence 20689, A
c	26	61.4	6.8	1026	7	US-10-496-351-17	Sequence 17, Appl
c	27	61.2	6.8	1419	11	US-11-096-568A-20533	Sequence 20533, A
c	28	61.2	6.8	1500	9	US-10-297-056-10	Sequence 10, Appl
c	29	60.8	6.7	1274	11	US-11-096-568A-21617	Sequence 21617, A
c	30	60.4	6.7	1453	11	US-11-096-568A-4228	Sequence 4228, Ap
c	31	60.2	6.7	1005	8	US-10-858-730-185	Sequence 185, App
c	32	60.2	6.7	1629	11	US-11-096-568A-13012	Sequence 13012, A
c	33	60.2	6.7	2247	7	US-10-496-351-8	Sequence 8, Appli
c	34	60	6.7	1540	11	US-11-096-568A-24562	Sequence 24562, A
c	35	59.8	6.6	1201	11	US-11-096-568A-25603	Sequence 25603, A
c	36	59.6	6.6	1392	11	US-11-096-568A-12509	Sequence 12509, A
c	37	59.2	6.6	11070	14	US-11-075-185-34	Sequence 34, Appl
c	38	59.2	6.6	78869	14	US-11-075-185-1	Sequence 1, Appli
	39	59	6.5	1047	7	US-10-496-351-11	Sequence 11, Appl
c	40	59	6.5	3513	8	US-10-858-730-142	Sequence 142, App
c	41	59	6.5	191684	14	US-11-121-086-2	Sequence 2, Appli
c	42	58.4	6.5	1614	11	US-11-096-568A-16842	Sequence 16842, A
c	43	58.4	6.5	6360	14	US-11-056-470-2	Sequence 2, Appli
c	44	58.2	6.5	1167	7	US-10-496-351-28	Sequence 28, Appl
c	45	58	6.4	1389	11	US-11-096-568A-23603	Sequence 23603, A

ALIGNMENTS

RESULT 1
US-10-496-351-2
; Sequence 2, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 9222
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-2

Query Match 8.7%; Score 78; DB 7; Length 9222;
Best Local Similarity 44.6%; Pred. No. 1.7e-08;
Matches 358; Conservative 0; Mismatches 435; Indels 9; Gaps 1;

```
Qy       66 GCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCC 125
           || |||| |   | ||  |   | || || ||   |  | ||| | |  | || ||   |
Db     7293 GCTGCTGGGCGGGGGCTTGGCGCTGGCCGCCGTCAACGGGCCGGGCGCGGTCGTGCTGTC 7352

Qy      126 GGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGC 185
            |  || || ||  ||||||   || ||  |||| |    || || ||          |
Db     7353 CGGTGACGGGGACGCGGTCGAGCGGGCCGCCGCCGAGTGCGCCGGGCGCGGGTGGCGGAC 7412

Qy      186 CGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGG 245
            | | |||     |       |   | || | || ||||  | ||||  ||        |
Db     7413 CAGGGCGCTGCGGGTGAGCCACGCCTTCCACTCCGCGCACCTGGACGGGGTGCTCGGTGA 7472

Qy      246 CCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGC 305
            || ||||| | || ||  ||  | |  |||| |  |||  |   |  ||  | | |
Db     7473 GCTGGCCGTGGCCGCGCGGCGGGTCGAGGTCCGGCCCCCGCGCGTCCCGCTGCTGTCCAC 7532
```

```
Qy        306 GGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCT 365
Db       7533 GCTCACCGGCGCGCCCGCCGACGTCGAGCTGCTCGCCGACCCGGACTACTGGGCTCGGCA 7592

Qy        366 CGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGC 425
Db       7593 GGCCAGGGAGGCCGTCCGGTTCGGCGACGGCGTCGCCCGGCTGGGCGAGCTGGGCGTGAC 7652

Qy        426 CCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCG---------AAGAAGC 476
Db       7653 GCGGGTGCTGGAGGTGGGGCCCGGCGGCGCGCTGGCCGCCCTGGTGCGCGAGCAGGTCGC 7712

Qy        477 GCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGA 536
Db       7713 GCCGGGGGCCGCGGTGGCGCGCGGGCTGGTCGACGGGGACGAGGTGGTCGGGCTGCTCTC 7772

Qy        537 CGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGC 596
Db       7773 CGGGATCGGCGACCTGTTCGCGTCCGGGGCCCCGGTCGACTGGACCGCGCTGCTGGGCGG 7832

Qy        597 GCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGA 656
Db       7833 GCCGGGGGCGGTCGTCGACCTGCCCACCTACCCGTTCCAGCGCACCAGGTTCTGGCCCGC 7892

Qy        657 GCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGG 716
Db       7893 GGCGCAGGCGCCCGGCGCGTGGTGCGAGGTCCGCCTGGAGCGGTTGCCCGCGCCTCGGGG 7952

Qy        717 TCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGG 776
Db       7953 CGGCGGGTCGGCGACCGTGCTGGAGGTGGGGCCGACCGCCTCGGCCGAGGAGCTGGCGAG 8012

Qy        777 TGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGA 836
Db       8013 CGTCGCCGACGTGCTGGCCGAGCACAGCGGCCCGCTGCTGGTGGTCACCGGGGACGGCGC 8072

Qy        837 TGTCGCGGGGGCGGAGCTCGCG 858
Db       8073 GGTGGCCGGGCTGGTGCGCTCG 8094
```

RESULT 2
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 82746
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

Query Match 8.7%; Score 78; DB 7; Length 82746;
Best Local Similarity 44.6%; Pred. No. 1.2e-08;
Matches 358; Conservative 0; Mismatches 435; Indels 9; Gaps 1;

```
Qy         66 GCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCC 125
Db      15670 GCTGCTGGGCGGGGGCTTGGCGCTGGCCGCCGTCAACGGGCCGGGCGCGGTCGTGCTGTC 15611

Qy        126 GGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGC 185
Db      15610 CGGTGACGGGGACGCGGTCGAGCGGGCCGCCGCCGAGTGCGCCGGGCGCGGGTGGCGGAC 15551
```

```
Qy        186 CGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGG 245
Db      15550 CAGGGCGCTGCGGGTGAGCCACGCCTTCCACTCCGCGCACCTGGACGGGGTGCTCGGTGA 15491

Qy        246 CCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGC 305
Db      15490 GCTGGCCGTGGCCGCGCGGCGGGTCGAGGTCCGGCCCCCGCGCGTCCCGCTGCTGTCCAC 15431

Qy        306 GGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCT 365
Db      15430 GCTCACCGGCGCGCCCGCCGACGTCGAGCTGCTCGCCGACCCGGACTACTGGGCTCGGCA 15371

Qy        366 CGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGC 425
Db      15370 GGCCAGGGAGGCCGTCCGGTTCGGCGACGGCGTCGCCCGGCTGGGCGAGCTGGGCGTGAC 15311

Qy        426 CCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCG---------AAGAAGC 476
Db      15310 GCGGGTGCTGGAGGTGGGGCCCGGCGGCGCGCTGGCCGCCCTGGTGCGCGAGCAGGTCGC 15251

Qy        477 GCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGA 536
Db      15250 GCCGGGGGCCGCGGTGGCGCGCGGGCTGGTCGACGGGGACGAGGTGGTCGGGCTGCTCTC 15191

Qy        537 CGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGC 596
Db      15190 CGGGATCGGCGACCTGTTCGCGTCCGGGGCCCCGGTCGACTGGACCGCGCTGCTGGGCGG 15131

Qy        597 GCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGA 656
Db      15130 GCCGGGGGCGGTCGTCGACCTGCCCACCTACCCGTTCCAGCGCACCAGGTTCTGGCCCGC 15071

Qy        657 GCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGG 716
Db      15070 GGCGCAGGCGCCCGGCGCGTGGTGCGAGGTCCGCCTGGAGCGGTTGCCCGCGCCTCGGGG 15011

Qy        717 TCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGG 776
Db      15010 CGGCGGGTCGGCGACCGTGCTGGAGGTGGGGCCGACCGCCTCGGCCGAGGAGCTGGCGAG 14951

Qy        777 TGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGA 836
Db      14950 CGTCGCCGACGTGCTGGCCGAGCACAGCGGCCCGCTGCTGGTGGTCACCGGGGACGGCGC 14891

Qy        837 TGTCGCGGGGGCGGAGCTCGCG 858
Db      14890 GGTGGCCGGGCTGGTGCGCTCG 14869
```

RESULT 3
US-10-496-351-2/c
; Sequence 2, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 9222
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-2

Query Match 8.6%; Score 77.8; DB 7; Length 9222;
Best Local Similarity 43.0%; Pred. No. 1.9e-08;

```
  Matches  382;  Conservative    0;  Mismatches  507;  Indels    0;  Gaps    0;

Qy         13 CCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGCCGCTG 72
Db       5820 CCGCACCGCGAGGTGGATCGCCACCAGCGAGGCCGAGCAGGCGGTGTCCACCGACAGCGC 5761

Qy         73 AGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGGCAGAG 132
Db       5760 GGGGCCCTCCAGCCCGAGCGCGTAGGCGATGCGCCCCGACACGACGCTGCCCGACACGCC 5701

Qy        133 GGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCGTGCCG 192
Db       5700 GGTGATCCGGTAGCCGTCGTCGCCGCGCGGGAGCACCTGGTCGTAGTCGTGGTTGGTGAC 5641

Qy        193 CGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCCTCGCC 252
Db       5640 GCCCGCGAACACGCCGACGTCCGTCCCCCGCACCGAGTCCGGGGCGATGCCCGCGCGCTC 5581

Qy        253 GTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGGATCAC 312
Db       5580 GAACGCCTCCCACGCGACTTCGAGCAGCAGCCGCTGCTGCGGGTCCATCGCGAGGGCCTC 5521

Qy        313 GCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCGCCGTA 372
Db       5520 GCGCGGGGAGATGCCGAAGAAGTCCGCGTCGAACCCGGCCGGGTCGTCCAGGAACGCGGC 5461

Qy        373 GAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCCGCGGC 432
Db       5460 CCGGTCGACGTAGGTGGAGCCCTCGTGGTCGGGGTCGGGGTGCACCAGCCGCTCCAGGTC 5401

Qy        433 CGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGC 492
Db       5400 CCAGCCCCGGTCGGCGGGGAACGGGCCGACCGCGTCGACGCCGCCCGCGACCAGGTCCCA 5341

Qy        493 GTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACAT 552
Db       5340 CAGGTCCTCGGGCGAGGCGACCCCGCCGGGCGCGCGGCAGGCCATGCCGACGATCGCGAT 5281

Qy        553 CGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCCCGGAA 612
Db       5280 CGGCTCGCCGGGGGCGGACGCGGGCGCGGGGGCGGACGCGCCGCCCGCCGGTCCCGAGGT 5221

Qy        613 CAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCGATCAC 672
Db       5220 CCCGCCCACCCGCTCCAGCAGCGCGGCGGCCAGCTCGACCGGGGTCGGGTGGTCGAACAC 5161

Qy        673 GAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACGATGTC 732
Db       5160 CAGCGTCGCGGGCAGCCGCAGCCCGGTGGCCTCGGCGAGCCGGTTGCGCAGCTCCACCCC 5101

Qy        733 CGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCGACGAG 792
Db       5100 GGTCAGCGAGTCGACCCCGAGGTCGCGGAACGCCCGCCGCTCGTCGACGCCCGCGACGGG 5041

Qy        793 CGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAG 852
Db       5040 CAGGGCGAGCACGGCGGCGGCCTCGGCCCGCACCAGCGCGAGCAGCGCGGCCGGGCCCCC 4981

Qy        853 CTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTCG 901
Db       4980 TGGTACCGGCGCGGCGGGCGCGGGCGCCCGGCGCGACGCGACCACCTCG 4932


RESULT 4
US-10-496-351-56
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
;   LENGTH: 82746
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

  Query Match             8.6%;  Score 77.8;  DB 7;  Length 82746;
  Best Local Similarity   43.0%;  Pred. No. 1.4e-08;
  Matches  382;  Conservative    0;  Mismatches  507;  Indels    0;  Gaps    0;

Qy         13 CCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGCCGCTG 72
Db      17143 CCGCACCGCGAGGTGGATCGCCACCAGCGAGGCCGAGCAGGCGGTGTCCACCGACAGCGC 17202

Qy         73 AGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGGCAGAG 132
Db      17203 GGGGCCCTCCAGCCCGAGCGCGTAGGCGATGCGCCCCGACACGACGCTGCCCGACACGCC 17262

Qy        133 GGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCGTGCCG 192
Db      17263 GGTGATCCGGTAGCCGTCGTCGCCGCGCGGGAGCACCTGGTCGTAGTCGTGGTTGGTGAC 17322

Qy        193 CGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCCTCGCC 252
Db      17323 GCCCGCGAACACGCCGACGTCCGTCCCCCGCACCGAGTCCGGGGCGATGCCCGCGCGCTC 17382

Qy        253 GTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGGATCAC 312
Db      17383 GAACGCCTCCCACGCGACTTCGAGCAGCAGCCGCTGCTGCGGGTCCATCGCGAGGGCCTC 17442

Qy        313 GCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCGCCGTA 372
Db      17443 GCGCGGGGAGATGCCGAAGAAGTCCGCGTCGAACCCGGCCGGGTCGTCCAGGAACGCGGC 17502

Qy        373 GAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCCGCGGC 432
Db      17503 CCGGTCGACGTAGGTGGAGCCCTCGTGGTCGGGGTCGGGGTGCACCAGCCGCTCCAGGTC 17562

Qy        433 CGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGC 492
Db      17563 CCAGCCCCGGTCGGCGGGGAACGGGCCGACCGCGTCGACGCCGCCCGCGACCAGGTCCCA 17622

Qy        493 GTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACAT 552
Db      17623 CAGGTCCTCGGGCGAGGCGACCCCGCCGGGCGCGCGGCAGGCCATGCCGACGATCGCGAT 17682

Qy        553 CGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCCCGGAA 612
Db      17683 CGGCTCGCCGGGGGCGGACGCGGGCGCGGGGGCGGACGCGCCGCCCGCCGGTCCCGAGGT 17742

Qy        613 CAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCGATCAC 672
Db      17743 CCCGCCCACCCGCTCCAGCAGCGCGGCGGCCAGCTCGACCGGGGTCGGGTGGTCGAACAC 17802

Qy        673 GAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACGATGTC 732
Db      17803 CAGCGTCGCGGGCAGCCGCAGCCCGGTGGCCTCGGCGAGCCGGTTGCGCAGCTCCACCCC 17862

Qy        733 CGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCGACGAG 792
Db      17863 GGTCAGCGAGTCGACCCCGAGGTCGCGGAACGCCCGCCGCTCGTCGACGCCCGCGACGGG 17922

Qy        793 CGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAG 852
Db      17923 CAGGGCGAGCACGGCGGCGGCCTCGGCCCGCACCAGCGCGAGCAGCGCGGCCGGGCCCCC 17982

Qy        853 CTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTCG 901
```

```
Db      17983 TGGTACCGGCGCGGCGGGCGCGGGCGCCCGGCGCGACGCGACCACCTCG 18031


RESULT 5
US-10-496-351-3/c
; Sequence 3, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 4770
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-3

  Query Match             8.4%;  Score 75.4;  DB 7;  Length 4770;
  Best Local Similarity   43.5%;  Pred. No. 7.3e-08;
  Matches  392;  Conservative    0;  Mismatches  506;  Indels    3;  Gaps    1;

Qy          2 GCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGC 61
              |||| | |     | | |    |  | |    |  | ||| ||||  | ||||  |     |
Db       3467 GCGGCGACGCCGCCGGTTCCGCCGGTGATCAGGACCGTGCCGCGCGGGGTCCACGGGGCG 3408

Qy         62 GCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGC 121
              || ||   ||| |||   ||| ||    |   |      || |   | |   |||  ||
Db       3407 GCCGGGGCCTGCGGTTCGGCGCGCACGAGCCTGCGCGCGAACACGCCCTGCGCGCGGATG 3348

Qy        122 GCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAG 181
              ||      |   |   | |||||||| | || | |||| |||  ||    ||     |
Db       3347 GCGACCTGGTCCTCACCTCCGTCGTCGACGCCGGGTCCGGCGAGGACGGCGGCGAGCCCG 3288

Qy        182 ATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGC 241
              ||  || |||| | |  | ||   |||   ||  || ||     |   |||| || |
Db       3287 CGGCGCTGGCGCGGGTCGGCGGGGTCGGCGAGGTCGACCAGCCCGCCCCAGCGGCCGGGG 3228

Qy        242 ACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGA 301
                ||    || ||||| |  ||      | |        | || | |   ||    |
Db       3227 TGCTCCAGCGCGACGACGCGGCCCAGTCCCCAGGTGGCGGCCTGGGCGGGGTCCGGCGGG 3168

Qy        302 TCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGC 361
              ||  |  |  |    || | ||  | |||  |  |||   | |  |  |||||  |
Db       3167 CCGTCGTGCGGGGAGATCGCGACGGCCCCGGTGGTGGCGCACCACAGCGGCGCGGTGGCG 3108

Qy        362 GGCTCGCCGTAGAGGGCGCTGTAC---GCGGCCACGAGCTCGTCCCATGTCGCCTCGCGG 418
                | || ||  ||| ||   |  |   | |  |||| |  || ||  || ||   | |
Db       3107 CCCGCGTCGCCGAGCGCCTGGACCAGGGTGAGCACGTCGGCGCCCGGCGTGGCAAGGACG 3048

Qy        419 CGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGC 478
              | |||| | ||||   ||| |  |  | |  || | |   |  |||  ||  |  ||
Db       3047 CCCGCGACGGCGGTGTGCGGGCGGAGCAGGTCGGCGAGGGCGGCGCGGTCGGCGTCGCCG 2988

Qy        479 CGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACG 538
              ||  || | |||||      |||  |  |      |  | | |    | || |||||
Db       2987 GGGTCGAGCGCGGCGCCGGGCAGCAGGTCGGCCAGCAGGTCGGGGTGCGGTCCGACGACG 2928

Qy        539 CGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGC 598
              ||      |   ||| ||    ||  | ||| | |  |  | ||||||     ||
Db       2927 AGCCAGTCGCCGCGCGGGGCGGACTCGGGCGGTTCGGGCAGGGTGCGCCAGGTGATCCGG 2868

Qy        599 ACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGC 658
              | ||     |  |||  |  |||  || |  |     |  |||
Db       2867 TAGCGCCAGGAGTCCAGCAGGGCCCGCTCGCTCCTGCGCTCGTGCCAGGCGGTCAGCGCG 2808

Qy        659 AGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTC 718
```

```
              || | ||      |  |  || || |  ||          ||  |       |  |
Db       2807 GGCAGGACGGCGCGCAGCGCGGCGCGCTGGTCGGGCTCGGTGAGCTCCAGCTCGGCGCAC 2748

Qy        719 GCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTG 778
              |    ||  |||      |   |  ||   |  | | |  ||  || |||   || |
Db       2747 AGGCCGTCGGCGTCGCGCGCCGCCGCCCACAGCGGGTCGTCGCCGCCCGCCTGCCGGGCG 2688

Qy        779 CCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATG 838
              || | ||| |   |  | | |  | |   |    ||| |  |  |  |  |    ||  |
Db       2687 TCGGGTTCGGCCTCCTGGAGCCAGTGGCGGGCGCGCTGGAACGGGTAGGTGGGCAGGTCG 2628

Qy        839 TCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCG 898
              ||||  ||  |  |  ||   ||  | || | || | |     ||||| || | | | ||
Db       2627 GCGCGCCGGGCGTCCAGGCCCTCGAAGACGGGGGCCCAGTCGACGGGCACGCCGCGCACG 2568

Qy        899 T 899
              |
Db       2567 T 2567


RESULT 6
US-10-478-943D-1/c
; Sequence 1, Application US/10478943D
; Publication No. US20060073574A1
; GENERAL INFORMATION:
;  APPLICANT: University College Dublin, National University of Ireland
;  TITLE OF INVENTION: Engineered biosynthesis of novel polyenes
;  FILE REFERENCE: PC-1636US
;  CURRENT APPLICATION NUMBER: US/10/478,943D
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: S2001/0527
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 113193
;   TYPE: DNA
;   ORGANISM: Streptomyces nodosus
US-10-478-943D-1

  Query Match             8.0%;  Score 72;  DB 7;  Length 113193;
  Best Local Similarity   44.2%;  Pred. No. 2.5e-07;
  Matches  392;  Conservative    0;  Mismatches  485;  Indels    9;  Gaps    2;

Qy         19 GGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGCCGCTGAGGTCG 78
              || ||| | |  |  |  | | | |  | | |  |   | | || || ||  | ||
Db      28552 GGCGGGCTGCTCCAGGAGCCACCGGCCGTAGCCGTCGGAGGCGGTGGCCCCCTCGGCACG 28493

Qy         79 CGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGGCAGAGGGCGAT 138
              || |   || | || || ||   ||||  |   ||   ||  |   | |||| ||| |
Db      28492 GTCGCGGGCCGCGCCGGCGAGGGCGGTCCTGGCCTCGGGCAGGCCGTCGAAGAGTGCGGT 28433

Qy        139 CCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCGTGCCGCGCGAT 198
              | |  |     || | |  | |||       |     |  ||  || | ||       | ||
Db      28432 CGGCCGGGTCTCGTGGTGCGCCGGGGCGAAGGTGTCCCAGCGGACGTCGGCGACGGTGAC 28373

Qy        199 GAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCCTCGCCGTCGAC 258
                |  |     || |  |  |       ||| |  |   ||  |||| ||| |||| |
Db      28372 CGCGGCGGAGCCGTCGGCGACGGCACCGGCGAGGGCGGTCAGGGCGGCGTCGGCGTCCAT 28313

Qy        259 GAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGGATCACGCGCCG 318
              | |  |     |   || ||| ||  |  | |    | |||   | | ||  ||
Db      28312 CACCGGCAGGCCGTTCATCCGCAGATGGCGGGTGGTCGACGGGTCGGTCAGCTCGGCCCA 28253

Qy        319 CGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCGCCGTAGAGGGC 378
              | |  ||  |  |||     || |  | ||| ||| |  || |  ||| |  |||||
Db      28252 GGCGTTCCAGGAGACGGCCAGCGCGGGCGTCCCCTCGGCGCGCAGGGCGCGGGCGAGGGC 28193

Qy        379 GCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCCGCGGCCGGCGC 438
              |     |||  |  | || ||      |  | || |||| |||||  | ||   ||
Db      28192 GTCCACATAGGCGCCTTCCGCGGCCTCGTCGGTGTGGCCGCGCACGCCCCAGACCCCGGC 28133

Qy        439 GTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGCGTATTC 498
```

```
Db      28132 GACG------GAGCCGAAGAGCACGAACGCGTCCAGGGAGCGGCCGGTGGCGGCCTCGGC 28079

Qy        499 GAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACATCGGATA 558

Db      28078 GAGGGTGTCCACCGGGGCGAGGGCGGCGGCCGGGCCGTCCGCCGGGGTGTCGTCGGCGTA 28019

Qy        559 ---GGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCCCGGAACAG 615

Db      28018 CACGACCGCGGTGAGGCCGGTCAGTCCGTCCAGGAGGGCGCGGGCCGCGTCCCGGTCGCC 27959

Qy        616 CGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCGATCACGAA 675

Db      27958 GCCGTCGTACGGGGCGACGGTGACCTCGGCGCCGAGCCGGCCGAGTTCGGCCCTCAGCTC 27899

Qy        676 GGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACGATGTCCGT 735

Db      27898 CTCGGTGTCGGTGGCACTGTCGGCCGGGGAGGTGGTCAGCACCAGGCGGGCGGCGCCCTC 27839

Qy        736 TTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCGACGAGCGC 795

Db      27838 GCGGGCGAGCCAGCGGGCGGCGCGGGCGCCCATGGCGCGGGTGCCGCCGACGACCAGGAC 27779

Qy        796 GCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAGCTC 855

Db      27778 GGTGCCGTGCGGGCGCCATCCGGAGTCGCCGGGGGCGACGGCGGTGCGCACCAGGCGGCG 27719

Qy        856 GCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTCG 901

Db      27718 GCCGAAGAGACCGGTGGCGCGGACGGCGATCTGGTCCTCGCCGCCG 27673


RESULT 7
US-10-496-351-1/c
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             8.0%;  Score 71.8;  DB 7;  Length 14055;
  Best Local Similarity   44.6%;  Pred. No. 3.8e-07;
  Matches  373;  Conservative    0;  Mismatches  457;  Indels    7;  Gaps    2;

Qy         48 GGGTGGAGATCGGCGCGGGCCGCTGAGGTCGCGCGGGCTACGGG-GCGACGCGGAGCGTC 106

Db       7970 GCGCCGAGTTCGGCGAGGTCGGCGGCCAGCTCGGCGGCGCCGGGCGCGTCGGGGCCGCGC 7911

Qy        107 CGCGAGGCGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGC 166

Db       7910 CGGGAGGTGATCAGCAGGTGGCGGGCGCCGTGCGCGGACACCAGGTGGCGGGCGGTGAGC 7851

Qy        167 TCGCGCACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGA 226

Db       7850 GCGCCGAGCGCGCCGCCGCCGTGGACCAGGACGGTGCCCTCGGGGTCGAGCGGGCGGGGG 7791

Qy        227 TCGACGCGGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGG 286

Db       7790 ACGGTGAGCACGACCTTGCCGACGTGCCCGCCGGACGCGACGTGCCGCAGCGCCTCGGGG 7731

Qy        287 ACCGACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTC 346

Db       7730 GCGTGCCGGACGTCGCGGACGGTCAGCGGCAGCGGCTCAAGGCGGCCCGCCGCGAACAGC 7671

Qy        347 CACACCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCAT 406

Db       7670 CCCAGCACCTCGGCGAGCAGTTCGGCGGTGCGGTCCGGGCCCGCGTCGCGCAGGTCGTAG 7611

Qy        407 GTCGCCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGC 466

Db       7610 GCC------CGGTAGGTCACGTCGGCGCGCACGAGGTCGCGCACGTCGACCTTGCCCAGC 7557

Qy        467 CCGAAGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAG 526

Db       7556 TCCAGGAACCGCCCGCCCGGCGCGAGCAGCCGCAGCGAGGCGTCCACCAGCTCGCCGGTG 7497

Qy        527 GCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGC 586

Db       7496 AGCGAGTTCAGCACGACGTCGACGTTCGGGAAGCGCTCGGCGAACCCGGTGTCGCGGGAG 7437

Qy        587 CAGAGATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTC 646

Db       7436 GAGGCCAGGTGCGCGTCGTCCAGGCCGAGCGCGCGCAGCACCGGCCACTTCGCCGGGGAC 7377

Qy        647 TTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTC 706

Db       7376 GCGGTGGCGAACACCTCCGCGCCCAGCAGCGCGGCCAGCCTGACGGCCGCGCCGCCGACG 7317

Qy        707 CTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGC 766

Db       7316 CCGCCCGCGCCCGCGTGCACCAGGACGCGGTCGCCCGGCCCGACCTGGGCGAGGTCGCGC 7257

Qy        767 GGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCG 826

Db       7256 AGGCCGTAGAGGGCGGTGAGGAACGCGACCGGCACCGAGGCGGCGCGCGCGAACCCCCAG 7197

Qy        827 AGGCCGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCG 883

Db       7196 CCGGGCGGGACGGGCGCGAGCAGCCGGGCGTCGGTGACCGCGACCGGCCCGAACGCG 7140


RESULT 8
US-11-096-568A-22025/c
; Sequence 22025, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;  TITLE OF INVENTION:  Therby
;  FILE REFERENCE: 2750-1592PUS2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE:  2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22025
;   LENGTH: 1773
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1773)
;   OTHER INFORMATION: Ceres Seq. ID no. 12407294
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1742)..(1742)
;   OTHER INFORMATION: n is a, c, t, g, unknown, or other
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1743)..(1743)
;   OTHER INFORMATION: n is a, c, t, g, unknown, or other
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1767)..(1767)
;   OTHER INFORMATION: n is a, c, t, g, unknown, or other
```

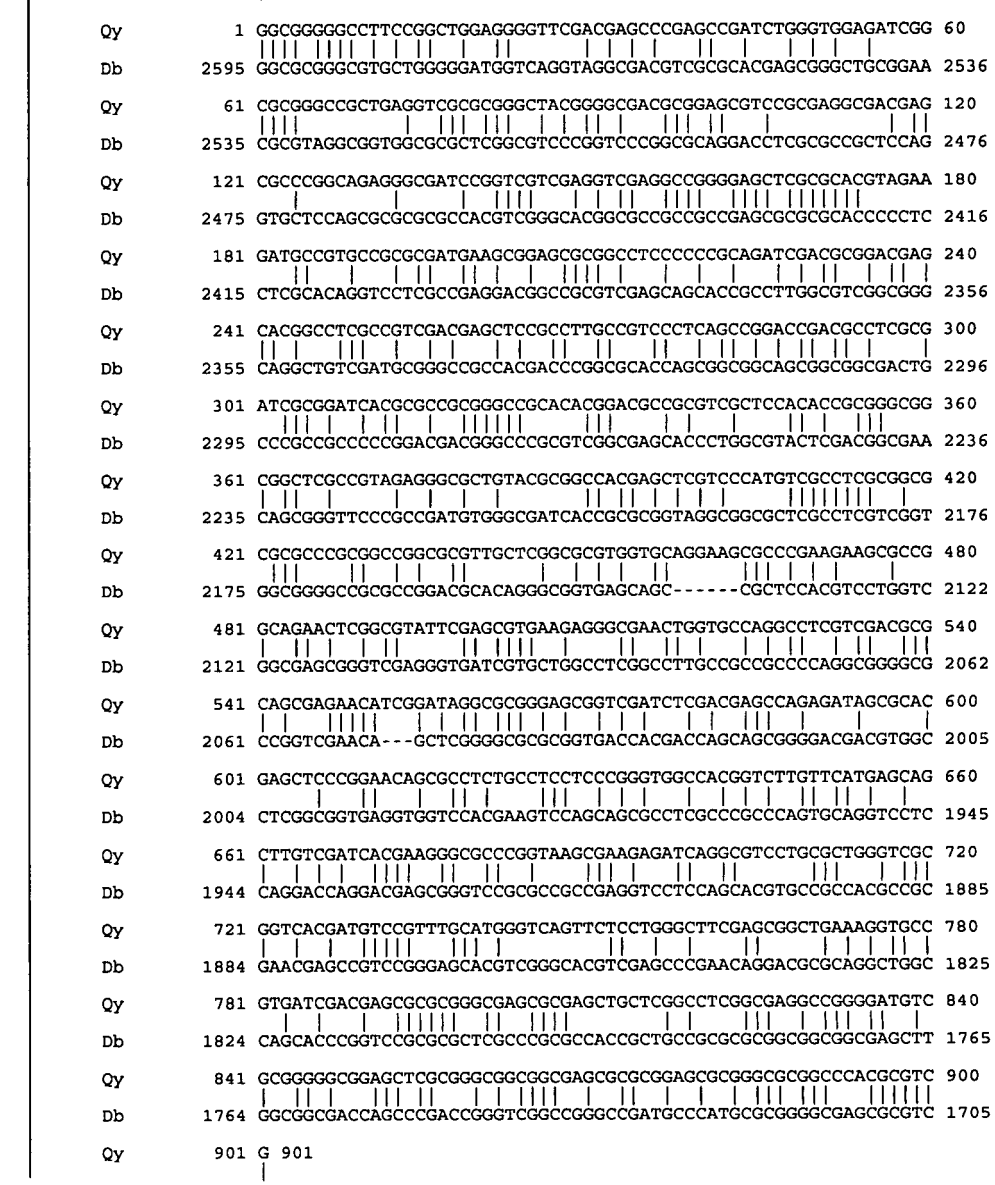
US-11-096-568A-22025

Query Match 7.9%; Score 71.6; DB 11; Length 1773;
Best Local Similarity 44.2%; Pred. No. 5.8e-07;
Matches 339; Conservative 0; Mismatches 425; Indels 3; Gaps 1;



RESULT 9
US-10-496-351-22/c
; Sequence 22, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 2742
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-22

Query Match 7.7%; Score 69.4; DB 7; Length 2742;
Best Local Similarity 44.0%; Pred. No. 1.7e-06;
Matches 396; Conservative 0; Mismatches 496; Indels 9; Gaps 2;



```
Db       1704 G 1704

RESULT 10
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT:  Zazopoulos, Emmanuel
;  APPLICANT:  Staffa, Alfredo
;  TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
;  FILE REFERENCE: 3002-2US
;  CURRENT APPLICATION NUMBER: US/11/205,109
;  CURRENT FILING DATE:  2005-08-17
;  PRIOR APPLICATION NUMBER: US/09/976,059
;  PRIOR FILING DATE: 2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 88421
;   TYPE: DNA
;   ORGANISM: Actinoplanes sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2077)..(3078)
;   OTHER INFORMATION: ORF 1; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3118)..(4032)
;   OTHER INFORMATION: ORF 2; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4038)..(5048)
;   OTHER INFORMATION: ORF 3; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6665)..(5814)
;   OTHER INFORMATION: ORF 4; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7703)..(6693)
;   OTHER INFORMATION: ORF 5; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9464)..(8130)
;   OTHER INFORMATION: ORF 6; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9691)..(10761)
;   OTHER INFORMATION: ORF 7; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (12751)..(10829)
;   OTHER INFORMATION: ORF 8; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13617)..(12802)
;   OTHER INFORMATION: ORF 9; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15203)..(13614)
;   OTHER INFORMATION: ORF 10; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15591)..(15863)
;   OTHER INFORMATION: ORF 11; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15880)..(19035)
;   OTHER INFORMATION: ORF 12; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19032)..(39713)
;   OTHER INFORMATION: ORF 13; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (39713)..(65800)
;   OTHER INFORMATION: ORF 14; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65826)..(66530)
;   OTHER INFORMATION: ORF 15; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (66546)..(67370)
;   OTHER INFORMATION: ORF 16; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (67384)..(70059)
;   OTHER INFORMATION: ORF 17; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70099)..(70662)
;   OTHER INFORMATION: ORF 18; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (70659)..(71906)
;   OTHER INFORMATION: ORF 19; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (73439)..(71964)
;   OTHER INFORMATION: ORF 20; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (74216)..(73563)
;   OTHER INFORMATION: ORF 21; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75424)..(74213)
;   OTHER INFORMATION: ORF 22; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75535)..(76464)
;   OTHER INFORMATION: ORF 23; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78110)..(76449)
;   OTHER INFORMATION: ORF 24; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79864)..(78107)
;   OTHER INFORMATION: ORF 25; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81624)..(79861)
;   OTHER INFORMATION: ORF 26; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (81909)..(81682)
;   OTHER INFORMATION: ORF 27; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82346)..(82062)
;   OTHER INFORMATION: ORF 28; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (82587)..(84446)
;   OTHER INFORMATION: ORF 29; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84481)..(85548)
;   OTHER INFORMATION: ORF 30; positive strandedness
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (85556)..(86845)
;   OTHER INFORMATION: ORF 31; positive strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87372)..(86803)
;   OTHER INFORMATION: ORF 32; negative strandedness
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87494)..(88420)
;   OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

  Query Match             7.7%;  Score 69;  DB 14;  Length 88421;
  Best Local Similarity   46.3%;  Pred. No. 1.2e-06;
  Matches  297;  Conservative    0;  Mismatches  340;  Indels    4;  Gaps    2;

Qy          5 GGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCG 64
              |||  |  ||||| |   ||| ||| ||  | ||     ||| |   | | ||| || |
Db      68265 GGGTGCGGCCGGCGGCGTGGGCTCGGCGGTCGCGCCGTCATCGGCCGGCGGCTCGCCGTG 68206

Qy         65 GGCCGCTGAGGTCGCGCGGGC-TACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGC 123
               ||||| |    |  |||||  |||  || ||  |  || |      ||| |||||
Db      68205 TGCCGCGGCCAGCATCCGGGCGAACGCTTCGGCGGCGGCCGGCGAGACCGCGCCGAGCCG 68146

Qy        124 CCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGAT 183
              |||| |       | |  |  | ||  | || |      ||     |  |||||  ||
Db      68145 GTGGCGGCACCGCAGCTCGGTGCCGCCGCCGGGCGTCCGGGAGACCTGCACGGTGAGCGG 68086

Qy        184 GCCGTGCCGCGCGATGAAGCG---GAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              |  |    |||| | |||| |   |  | || ||  |       || ||| ||  || |
Db      68085 GTGGGTGGGCGCCAGGAAGGGCACGTACTCGTCCGGCGGGGCGAGGTCGCCGTCGAAGGC 68026

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              ||  ||| |   | ||  || | | ||   ||||   | |      | |   |  |
Db      68025 CAGCGCCACCGGGCCGGGGACCGCGGCGGCGCCGGAGCCCGAGGCGGCGCTGGAGCCCGC 67966

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              |      | || |  ||   ||| || |  ||||||  ||| | |  |||| | ||
Db      67965 CGCCACACCCGAGCCCGCCGCCGCGCCCAAGCCCGCCGCCGCGCCCGAGCCCGCAGCCGC 67906

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              |   ||||  ||||| |  |  ||| |      ||  |||   ||||  |  |
Db      67905 GCCCGGCTCGTACCGGGCGAGGACCTTGGCGAGCGCGGCGGGCCAGCGCGCCGGGTCGGC 67846

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
              || |||    |||||| ||      |  ||     |   |    | |||  |  || | |
Db      67845 CGGGCCGCGCGCCGGCACGCGCAGCTCGGCCCACCGGTCGTCCGGGCCGCCGTCGCCCAG 67786

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
              ||   |  |||       |       | ||| | | | |   |  | |    ||||
Db      67785 GCCCCAGGCGGGCGCGGCCGAGGGCGGCGGCGCGTCCGGTTCCGCCGGTTTGCGCCCGCG 67726

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              | |||    |  |||   |||  ||| |          ||| |  |  | | | |||
Db      67725 CGGCGCCCGCTCCGGGCGGGCCAGGGCGATCAGCGCGTCGCGGTCCAGGCGGTCGCGGCG 67666

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCA 641
              |  | || | | |||  |  | |   | |||  |  ||| |
Db      67665 GGCCACCAGCACCAGGTCGGCCGAGCCGTCCGCGTAGGCGA 67625


RESULT 11
US-11-096-568A-27264/c
; Sequence 27264, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;  TITLE OF INVENTION:  Therby
;  FILE REFERENCE: 2750-1592PUS2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE:  2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 27264
;   LENGTH: 1339
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1339)
;   OTHER INFORMATION: Ceres Seq. ID no. 15180690
US-11-096-568A-27264

  Query Match             7.6%;  Score 68.4;  DB 11;  Length 1339;
  Best Local Similarity   46.0%;  Pred. No. 3.1e-06;
  Matches  231;  Conservative    0;  Mismatches  271;  Indels    0;  Gaps    0;

Qy        100 GAGCGTCCGCGAGGCGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCC 159
              ||||    | ||| | | |     |   || |||  |||     | |   |||  ||  ||
Db        605 GAGCACGAGGGAGCCCAGGTTGACGGCGCCGAGTACGACGGCGCCCATGAGCGCGGCGCT 546

Qy        160 GGGGAGCTCGCGCACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCC 219
              |          ||  |||     ||   || || |||||  ||||  ||||  ||    | ||
Db        545 GCTCCCGAAGCCGACGGTCCGGAACACGAGCGGCGCGAAGAAGGCGAGCACGATGACGCC 486

Qy        220 CCGCAGATCGACGCGGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCT 279
              |   || | || |   | | ||||||| || ||   || || | ||     |   || |
Db        485 CGTGAGCTGGAAGAACATGGGCACGGCCACGGCGAGCACCAGGTGCGGGCGGTACTCGCG 426

Qy        280 CAGCCGGACCGACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGC 339
              | ||  | ||  |      |    ||| |  | |||  ||  |  ||  | |    |||
Db        425 CCGCGTGGCCATCCGGCGGAACGCGCCGTCCTCGCCGCGGCGCGCGGCCTCCACGGCCTT 366

Qy        340 GTCGCTCCACACCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTC 399
              | || |   |          | |  | || ||  |  ||  | |  |    |  | ||| |
Db        365 GGCGATGTCCCTGAGCTCGGCCTCCACGTCGGCGTCGGGCCCGCGCACCCGGAGCAGCGC 306

Qy        400 GTCCCATGTCGCCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAG 459
              |||  | | | |||  |||| | | |     | |   | ||   |  ||  | | |
Db        305 CGCCCGGGCCCCGTCGGCGCGCCCCCGCATCACGAGGCTGCTGGGGGTGTCGGTGAGGAA 246

Qy        460 GAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTG 519
              ||      |||| ||||      | |  |   |||     | ||    |    | ||  |
Db        245 GAGCGCGCCGACGAAGATGGCGACGGCCGGCGCGCCCGCCAGGCCCAGCGAGACGCGCCA 186

Qy        520 GTGCCAGGCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTC 579
              |||||   |||   |  |     ||  ||   | | ||||           ||| | |
Db        185 CCCCCAGGAGGCGTGCGCCGTCGCGTAGTTCACGAGGTTGGCGATCAGCACCCCGAGCGC 126

Qy        580 GACGAGCCAGAGATAGCGCACG 601
              || ||    |  | |||| | ||
Db        125 GAGGAAGAACTGGTAGCCCGCG 104


RESULT 12
US-10-496-351-4/c
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 9975
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-4
```

```
  Query Match             7.5%;  Score 68;  DB 7;  Length 9975;
  Best Local Similarity   45.1%;  Pred. No. 2.8e-06;
  Matches  386;  Conservative    0;  Mismatches  455;  Indels   15;  Gaps    3;

Qy         49 GGTGGAGATCGGCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCG 108
              ||||  |  |    |||  |||| |   ||  |     | |  ||||  |     |
Db       9882 GGTGACGCGCACGCCGGTGCGCTCGGCCAGCAGGTTGCGCAGCTCGACGGCGGTGAGCGA 9823

Qy        109 CGAGGCGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTC 168
               | |   ||   ||| | || |  ||  | ||||| ||   ||  | ||  || |  |
Db       9822 GGTGAACCCGGTGTCCCAGAAGGGCTCGTCCGGGTCGACGGACTCAGGCCCCCGGTGGCC 9763

Qy        169 GCGCACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATC 228
              | | |||  |  | || ||   ||| |||      | |||    ||  |  |    | ||
Db       9762 GAGGACGGTGGCGGTGGCGGCGCGCACGAGCGTCAGCAGCAGCTCCGGCAGCTCGTGGTC 9703

Qy        229 GACGCGGACGAGCA------CGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAG 282
              | ||    || |||      || ||  | |   | ||| ||| ||   | ||  |||  |
Db       9702 GGCGGCCCCGCGCAGCCGATCGACCGGGTCCGGGGCGACCTCGGCGCGGGCGCTCCTGCG 9643

Qy        283 CCGGACCGACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTC 342
              ||| || | | |     |  | | |  ||      ||  |   | || |||| |
Db       9642 CCGCACGGGCACGAGCTGGCGCAGCAGCGGCGGCACGTCCGCGTCCTGCGGCGTCGCGGC 9583

Qy        343 GCTCCACACCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTC 402
               | || ||  | | || ||  || |  |||| |   ||| | |  ||  ||
Db       9582 CAGGCTCAGCGGCGCCAGCAGCGGGCGGCCGTCGGCGAGGGCGGCGTCGAACAGGGCG-- 9525

Qy        403 CCATGTCGCCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAA 462
              || |||||  |     |  || |  | || ||  ||  |  |||| |  |  |
Db       9524 ----GTGGCCTCGGCGTCGTCCAGCGGGTGCAGCCCGGCGCGGGAGGCGTCGACCCCGGC 9469

Qy        463 GCGCCCGAAGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTG 522
              |  |  |  | ||||||       |  |||      |||||  |   ||  |   |    |
Db       9468 GGCCATGCCGGAGCGCCACGGCCCCCAGGCCAGGGAGAGCGCGTGCAGACCCGCCGCCCG 9409

Qy        523 CCAGGCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGAC 582
              ||       || | | ||    |  |||          ||   | ||  | |     ||
Db       9408 CCGTCGTGCGGCCAGGCCGTCCAGGAAGGCGTTCGCCGCCGCGTAGTTGGCCTGGCCGGG 9349

Qy        583 GAGCCAGAGATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCAC 642
              |   | |||   || | |  |     |||||| |     |  |    |  | | |
Db       9348 GGTGCCGAGCACGCCCGCCGCCGAGGAGAACAGCACGAACAGCGTCGTGGTGCCGTCGCG 9289

Qy        643 GGT---CTTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATC 699
              |||   || || || | |        | | | |  ||||||| |    |||   ||
Db       9288 GGTCAGCTCGTCCAGGTGGGCGGCCACCAGCGCCTTGGGCGCCAGCACGGCGTCCAGCCG 9229

Qy        700 AGGCGTCCTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGC 759
              |  |   || |  || |    |  ||  ||||          |   || | |  | |
Db       9228 GGCGGGGGTGAGGGCGGACAGGACGCCGTCGTCCAGGACGCCCGCGGTGTGCACGACGCC 9169

Qy        760 TTCGAGCGGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGC 819
               || |  | ||  || |    |  || ||||||||     |  || ||| |  ||||
Db       9168 GCCGGGTCGGTGCTCGGCGAGCAGCGCGGCGAGCGCGGCCCGGTCCGTGAGGTCGGCGGC 9109

Qy        820 CTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAG 879
              |   | || || |   | || |      ||| |||  |||     || ||||||  | |
Db       9108 GACCAGGTGGGCGCGCGCGCCGAGTCCCGCGAGGTCGGCGGCCAGTTCGGGCGCGCCGGG 9049

Qy        880 CGCGGGCGCGGCCCAC 895
              |||||  | | | | |
Db       9048 CGCGGCGGGGCCGCGC 9033


RESULT 13
US-10-496-351-12/c
; Sequence 12, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 3354
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-12

  Query Match             7.5%;  Score 67.6;  DB 7;  Length 3354;
  Best Local Similarity   45.7%;  Pred. No. 4e-06;
  Matches  411;  Conservative    0;  Mismatches  464;  Indels   25;  Gaps    4;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              ||||| | |||| ||   ||||  | |     ||  | | ||  || | || |   |
Db       3192 GGCGGCGACCTTGCGCTCGGAGCCGGTGAGGCCGCTGGTGTCGTGCTCGACGGGGTCGGC 3133

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
              ||||  | | |   | |||||| |  |||      |||    || ||  |     |
Db       3132 CGCGTCCGGTTCCAGGCGCGCGCACAGCGGTCGCGCGCCACACGCCCTGGCCAGGTGCCA 3073

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              ||||  ||  ||  ||   ||| ||  | ||| |  |||| ||    | | ||  ||  |
Db       3072 GGCCCTGCGCAGCACGGCGCGGGCGCGGCGGTGGTCGCCGAGGGCGTGGTGCGCGGCGCC 3013

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              || || |    | | |  |          ||   |||    ||   | ||  |||  | ||
Db       3012 GAGGCAGGCGAGGGTGGCG---------GCCAGCTCGTAGCGGTCACCGCAGCGCTCCAG 2962

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              || |  || | ||| ||||      |  || |    || || ||    ||  | | |
Db       2961 GGCGTCGGCGGCCTCGGCGAGCAGGTGCGGGCGGCGGGTCGGCGGGCTGAGCGGGGCGAG 2902

Qy        301 ATCGCGGATCAC----------GCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCA 348
              ||||  | | |           ||| | || | ||    ||| ||  |||
Db       2901 CGCGCGCAGCGCCAGCGCCCTGGGGCGGCCCGCGTCGGCGCCGGTCGTGGCGAGCTGGTC 2842

Qy        349 CACCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGT 408
              | | |  || || |  ||| || | ||  ||        || |     |||| |
Db       2841 GAACACCAGCCGCCGGGCGCGGTCGTGGTTGCCCAGGCGGGTCCACGCCTCGGCGGCCCC 2782

Qy        409 CGCCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCC 468
              |     |  |||||    ||| | |  ||  || | || | | | |||  ||  |
Db       2781 GGTGCGCCACGGCGCGAGGCCGGGCAGGTCGAGGCCCCACGAGCGCAGGAGGTCGCCGCA 2722

Qy        469 GAAGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGC 528
              |  |      || | | ||||  ||  ||  | |   |  |  |||| | |
Db       2721 CGCCAGGAAGTCCGCGAGCGCCGCGTGGTCCCGCCCGGACGCGAGGTGGTGGTGGCCGCG 2662

Qy        529 CTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCA 588
              | |||  | |    ||| |       ||   | ||  |  |  |||   | ||
Db       2661 CGCGTGCAGGTAGTGCGGGGCGGCGGGGCCGTCGAAGAGGGCGTCCGGCACGGGCGCGGT 2602

Qy        589 GAGATAGCGCACGAGCTCCCGGAACAGCGC---CTCTGCCTCCTCCCGGGTGGCCACGGT 645
              ||  |  ||| ||  |||||  | || |    |  |  || |  || |  |  |  |
Db       2601 GACGTGCCGCCCGGCCTCCCGGCGCCGCCCGCGCAGGGTGTCGGCGAGGATCAGCGAGCC 2542

Qy        646 CTTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGT 705
              |     || |  ||     ||  |  | | |||   ||   |||  |||  | |
Db       2541 CAGCGGCAGGCCCACGGACGCGCCCCAGCCGTGCGAGGGGACGGCGTCGAGCGCGGCGCG 2482

Qy        706 CCTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTC-GA 764
                | ||  ||  |||  | |||  |||      ||      |   | || ||  |||  | |
Db       2481 GGCGTGCCCGGCGGCGCCCGCGAGGTCGCCGAGCCGCAGCGCGATCACCGAGGCGGCGGC 2422

Qy        765 GCGGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGG 824
              |  |  | | |   ||  | | | ||  ||||  ||    |  |  |  |  | |||  |
Db       2421 GGCGAAGGAGGCGACCCAGGTGGGCGCGCGCCGGGCGCGGGCCTCGTCGAGCAGCCGGGC 2362
```

```
Qy        825 CGAGGCCGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGG 884
              ||||| ||   |   ||| |||| |   ||| ||  | | |   ||  || | ||| |
Db       2361 GCAGGCCCGGTCGGCGTCGGCGGCGCGGTCGGCGCGCACCAGGGCTAGCAGGGCGAGCAG 2302


RESULT 14
US-10-496-351-57/c
; Sequence 57, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
;   LENGTH: 11905
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-57

  Query Match             7.5%;  Score 67.4;  DB 7;  Length 11905;
  Best Local Similarity   45.4%;  Pred. No. 3.7e-06;
  Matches  413;  Conservative    0;  Mismatches  476;  Indels   20;  Gaps    4;

Qy          2 GCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGC 61
              | || |  |  ||  | ||   ||  |||   | ||||||    | |   || |  ||||
Db       9909 GTGGTGCTCGGCCACCCGGTCCGGGCCGATCCGGCCGAGCGCGCCCGCCAGGCGGCCGGC 9850

Qy         62 GCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGC 121
              |  | |||     | |||||  | |   ||  | | |||   |||     |  || | |||||
Db       9849 GAAGACCGTGCCGATCGCGACGCCCTCGCCGTGCCGCACCGCGAAACC--GGTGGCGAGC 9792

Qy        122 GCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAG 181
              || ||  | || |||     || | ||  || ||||  | |  ||  ||     || |
Db       9791 TCCAGGGCGTGGCCGAGGGTGTGGCCGTAGTTGAGGGTGTGCCGCAGGCCGCGGTCGCGC 9732

Qy        182 ATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGC 241
              | ||     | | |  |  |   |||  | | |  |    |   |    |  |||
Db       9731 TCGTCGGCGGCCACCACGCGGGCCTTGAGGGCGACGCTCGCCGCCACCTGGTCCAGCAGC 9672

Qy        242 ACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGA 301
              |  | |   | |  || |  |  ||  ||  |   |  || |  |    ||
Db       9671 GGCAACCGGTCAAGGCCCGGCGCGCCGATGAAGTGGCAGCGGGCGATCTCGCCGAGGCCG 9612

Qy        302 TCGCGGATCACGCGCCGCGGGCCGCACACGGAC---GCCGCGTCGCTCCACACCGCGGGC 358
              | ||| | | |||||   ||    |    ||  |    |   |||| |  ||| ||||
Db       9611 TTGCGCAGCTCGCGCTCGGGCAGGGTGGCGAGCAGGTCGAGGTCGCACAGCACGGCGGCG 9552

Qy        359 GGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGG 418
              ||| ||  |  | ||  | ||  || |  | || || |   ||     | || || || |
Db       9551 GGCTGCCAGTAGGAGCCGACGAGGTTCTTGCCCTCGGGGAGGTTGACCGCCGTCTTGCCG 9492

Qy        419 CGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGC 478
              |  ||| |||| |          |  | | ||| |  |  |    || ||    | |  |
Db       9491 CCGACGCTCGCGTC---------GACCTGGGCGAGCAGCGAGGTCGGCACGTGCACGACC 9441

Qy        479 CGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACG 538
              |   |  |  |  |||        | |  |||| ||| |  ||    ||     | || ||
Db       9440 GGTGTGCCCCGGTGGTAGAGCGAGGCGGCGAGGCCGACCGCGTCGGTGGTGGTGCCGCCG 9381

Qy        539 CGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGC------CAGAGA 592
              |    ||||  |     | ||||||||   || ||| |||| ||| |      || ||
Db       9380 CCGCAGGAGACGACGACGTCGGCGCGGGTCAGGCCGAACTCGGCGAACCGGCTGCACAGG 9321

Qy        593 TAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTC 652
              | |  |||     | ||  |   | |   |  | || |||| |  | |  |      |
Db       9320 TGGGCGACGGTGGCGAGGGTCTTGTCGTGCTCGCCGTCGCGGGCCGGGAGGACGAGGGAG 9261

Qy        653 ATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGC 712
              ||          ||||  |     |  || | ||   |||  ||   |    | | ||||
Db       9260 GGGACGCCGGGGTCGGGCGTCTGGTCCGGCGGGCGGGCGGTGACCACGACGGCGCGGCGC 9201

Qy        713 TGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGA 772
              |      || |   | | |     | | ||    |  | |   |  | |||   |
Db       9200 GCGCCGAGGGCCCGCACGACGTCCGGGAGGGCGGCGCGCACGCCGTGTCCGATGTGGACG 9141

Qy        773 AAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCG 832
              |  | || | ||| | ||| ||      || | || |       |||   || | |  |
Db       9140 GTGTAGGCGCGCTCGCCCAGCTCGACCCGGACCTCGCGGGTGGTGGCGGTGGTGGGCACT 9081

Qy        833 GGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCC 892
              |     ||  |||  || |    ||  |  ||    |  |||||||| |||||||| |
Db       9080 GCTTCCTCCTCGGGAGGGCGGGACGGGTGGTGGGGCGGGCGGCGGAGCGAGGGCGCGGGC 9021

Qy        893 CACGCGTCG 901
              |    | ||
Db       9020 CCTCAGCCG 9012


RESULT 15
US-11-017-550-68/c
; Sequence 68, Application US/11017550
; Publication No. US20050250183A1
; GENERAL INFORMATION:
;  APPLICANT: The Scripps Research Institute
;  APPLICANT:  Schultz, Peter G
;  APPLICANT:  Wang, Lei
;  APPLICANT:  Anderson, John C
;  APPLICANT:  Chin, Jason
;  APPLICANT:  Liu, David R
;  APPLICANT:  Magliery, Thomas
;  APPLICANT:  Meggers, Eric L
;  APPLICANT:  Mehl, Ryan A
;  APPLICANT:  Pastrnak, Miro
;  APPLICANT:  Santoro, Stephen  W
;  APPLICANT:  Zhang, Zhiwen
;  TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
;  FILE REFERENCE: 54-000120US
;  CURRENT APPLICATION NUMBER: US/11/017,550
;  CURRENT FILING DATE:  2004-12-17
;  PRIOR APPLICATION NUMBER: US/10/126,927
;  PRIOR FILING DATE: 2002-04-19
;  PRIOR APPLICATION NUMBER: US 60/285,030
;  PRIOR FILING DATE: 2001-04-19
;  PRIOR APPLICATION NUMBER: US 60/355,514
;  PRIOR FILING DATE: 2002-02-06
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68
;   LENGTH: 3305
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-11-017-550-68

  Query Match             7.3%;  Score 65.8;  DB 14;  Length 3305;
  Best Local Similarity   43.5%;  Pred. No. 1e-05;
  Matches  350;  Conservative    0;  Mismatches  452;  Indels    3;  Gaps    1;

Qy         91 GGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAG 150
              | || ||| || |||  |  |     || |||    || |||  |   | | |||  ||
Db       1422 GTCGGCGCCGATCGTGAGGAACCGCTCGGGCGAGGCGCTGAGCACCGACAGCTCGGGGAA 1363

Qy        151 GTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGC 210
              |||||   || |       | |||  |  ||||| | || |||||  | || |  ||
Db       1362 CTCGAGCAGGGCGCCAGACGGGACGGGGCTGATGCGGCGCAGCGCGGAGTAGAGCGGCAG 1303

Qy        211 GGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCCTCGCCG---TCGACGAGCTCCGC 267
              ||||  | ||           ||  ||| |   || |  || |   |||     ||| |
Db       1302 GGCCGTCGCCTCGGTCGGCGCGGTGACCATGTTGGTCAGGCAGATCTCGTACGACTCGCC 1243
```

```
Qy        268 CTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCA 327
              |||| |  |  |     | ||    | | |||  ||||   || | | |    ||||
Db       1242 GTTGCGGAGCGCCGAGGCGTCCTTGTCGTGGCGTGCGCGAGCCAGGGGGCCGAAGCCGGC 1183

Qy        328 CACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGC 387
              | | | | | | | | |   | |||  ||  || | |   |       || ||| |   ||
Db       1182 CGCCGCCTCGGGGACCCCGAAGACCATGGCGGGGGTCGGCCTCGGCCGGACGCGGACGGC 1123

Qy        388 GGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGG 447
                ||      |  |    | || ||| ||  ||  | |   |||      ||| |
Db       1122 CAGGCCGGTGAGGGTCTCGGCCGTCTCCCGCAGCCAGGCGCGGGCGCCGTCGTCGTGGCC 1063

Qy        448 CGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAA 507
              |  | |||  |||    ||  | || |||  |        | | ||    |||  | |
Db       1062 CCGGCGGTCGAGGGCCAGCAGGTAGCAGCAGCCTTCCTGGTGGTCGAGGGCGATGGCGCG 1003

Qy        508 GAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGA 567
              |  ||||||  ||  |  ||| |||  |    |   ||  |     || | | ||  ||
Db       1002 GTCGGCGAAGAGGAACGCGGCGTCGGGGTGCGGGGCCGGTACTGCGGGGTCGCCGGTGGT 943

Qy        568 GCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTC 627
                |  ||    ||   | ||| ||||  |||      |  ||   | | |    | |
Db        942 CTCCGCCTTCAGCTCGTAGCCGAGGTAGCCGACGTAGCCGAGGTTGAACTCGAAGGGCAG 883

Qy        628 CTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTA 687
              ||  ||| | |      |      |      ||||  ||   || | ||       ||
Db        882 GTCGGGGGCGACGGGACCCGCCGGCGGTTCGAGCTGCTCCTCCAGGTAGCTGAAGAGGGT 823

Qy        688 AGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCA 747
              ||   |      | |||  |||  | ||   ||   |||  | |          ||
Db        822 CGCCGCGTCCCGGGTCGTGGTGCCGTCGGAGCCGCGGACGGAGACGACGCCGTCGGCGAC 763

Qy        748 GTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGC 807
              |   |   |   |    |   |   |  || ||  ||| ||||   |  |    |||
Db        762 GCGGTAGGTGAGGTACTCGGCGAGCGGGCCGCGGTCGTCGCCGAGGAAGGAGAACGGCGA 703

Qy        808 GAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGC 867
              |      |||    ||| |  || |   |    |   | |  |  ||||   | |  |||
Db        702 GGCGCCTTCGAGGACGGAGCTGCTGTCCAGCCAGAACGTGGCGCCCTCGGCCGGGCAGGC 643

Qy        868 GAGCGCGCGGAGCGCGGGCGCGGCC 892
                  ||| | | |    ||| ||
Db        642 AGCGCGGCGTACCTCTTCGGCGTCC 618
```

Search completed: April 27, 2006, 15:53:02
Job time : 566.78 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 4465.48 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_76000_76900
Perfect score: 901
Sequence: 1 ggcgggggccttccggctgg.....cgggcgcggcccacgcgtcg 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

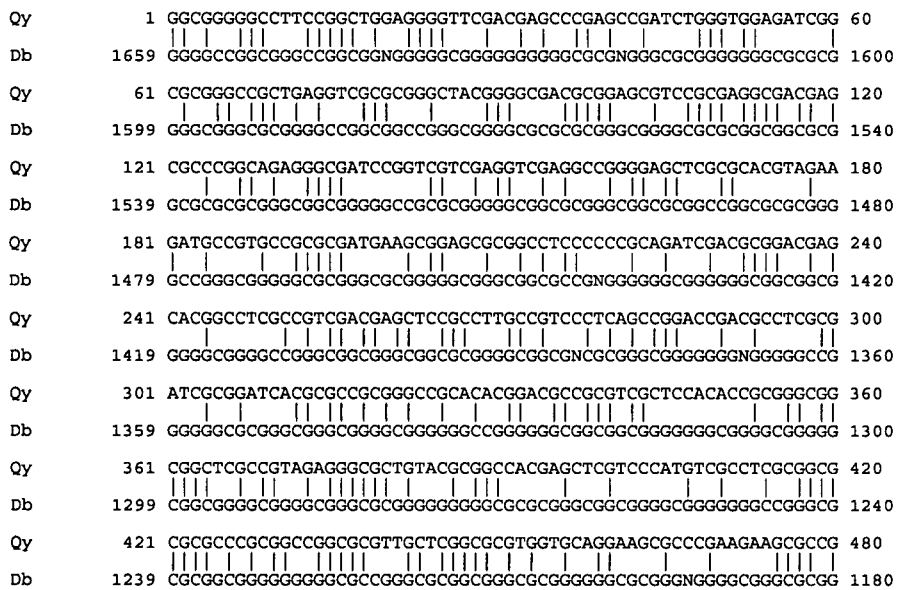
SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	97.2	10.8	1674	10	CL078342	CL078342 CH216-149
	2	95	10.5	1309	10	AG077201	AG077201 Pan trogl
c	3	94.4	10.5	2332	10	AG363333	AG363333 Mus muscu
c	4	93.8	10.4	1713	10	AG435689	AG435689 Mus muscu
	5	93.8	10.4	1822	10	AG435170	AG435170 Mus muscu
	6	93.4	10.4	1280	10	CL470317	CL470317 SAIL_140_
c	7	92	10.2	1413	10	CL505027	CL505027 SAIL_744_
c	8	91.8	10.2	1694	10	CL509291	CL509291 SAIL_811_
	9	91.6	10.2	1798	10	AG171124	AG171124 Pan trogl
c	10	90.2	10.0	1288	5	BQ678719	BQ678719 AGENCOURT
	11	89.2	9.9	1462	10	AG441877	AG441877 Mus muscu
	12	88.6	9.8	1516	2	BG809984	BG809984 mgct002xd
	13	88.6	9.8	1523	10	AG448267	AG448267 Mus muscu
	14	88.4	9.8	932	10	CNS0072Q	AL066742 Drosophil
c	15	88.2	9.8	1780	10	AG448243	AG448243 Mus muscu
c	16	88	9.8	1610	9	BZ569386	BZ569386 pacs2-164
	17	88	9.8	1628	10	CG757066	CG757066 P052-2-A0
	18	87	9.7	2332	10	AG363333	AG363333 Mus muscu
	19	86.6	9.6	1878	10	AG393083	AG393083 Mus muscu
c	20	86.4	9.6	1885	10	AG363563	AG363563 Mus muscu
	21	86.2	9.6	1440	10	AG332984	AG332984 Mus muscu
c	22	85.8	9.5	932	10	CNS0072Q	AL066742 Drosophil
	23	85.6	9.5	1116	9	BZ569478	BZ569478 pacs2-164
	24	85.6	9.5	1713	10	AG435689	AG435689 Mus muscu
c	25	85.6	9.5	2041	10	AG363808	AG363808 Mus muscu
c	26	85.4	9.5	1454	10	AG290167	AG290167 Mus muscu
	27	85.4	9.5	1552	10	AG430101	AG430101 Mus muscu
c	28	85.2	9.5	1061	10	AG046200	AG046200 Pan trogl
	29	85.2	9.5	1267	10	CL476590	CL476590 SAIL_258_
	30	85.2	9.5	1421	9	BZ569488	BZ569488 pacs2-164
c	31	84.8	9.4	1406	10	CG756569	CG756569 P051-4-B0
c	32	84.8	9.4	1542	10	AG032943	AG032943 Pan trogl
	33	84	9.3	1299	10	AG039481	AG039481 Pan trogl
	34	83.8	9.3	1885	10	AG363563	AG363563 Mus muscu
c	35	83.6	9.3	1785	9	CC219595	CC219595 CH261-62D
	36	83.4	9.3	935	10	CNS006XK	AL066051 Drosophil
c	37	83.2	9.2	1134	10	AG043044	AG043044 Pan trogl
c	38	83.2	9.2	1281	2	BG852363	BG852363 1024034A0
c	39	83	9.2	1625	10	AG043477	AG043477 Pan trogl
c	40	83	9.2	1798	10	AG171124	AG171124 Pan trogl
	41	82.8	9.2	1137	2	BG809979	BG809979 mgct002xd
c	42	82.6	9.2	1101	10	AG039543	AG039543 Pan trogl
	43	82.4	9.1	1772	10	CL508812	CL508812 SAIL_802_
c	44	82.2	9.1	1695	9	CC290874	CC290874 CH261-172
	45	82	9.1	1692	10	AG396765	AG396765 Mus muscu

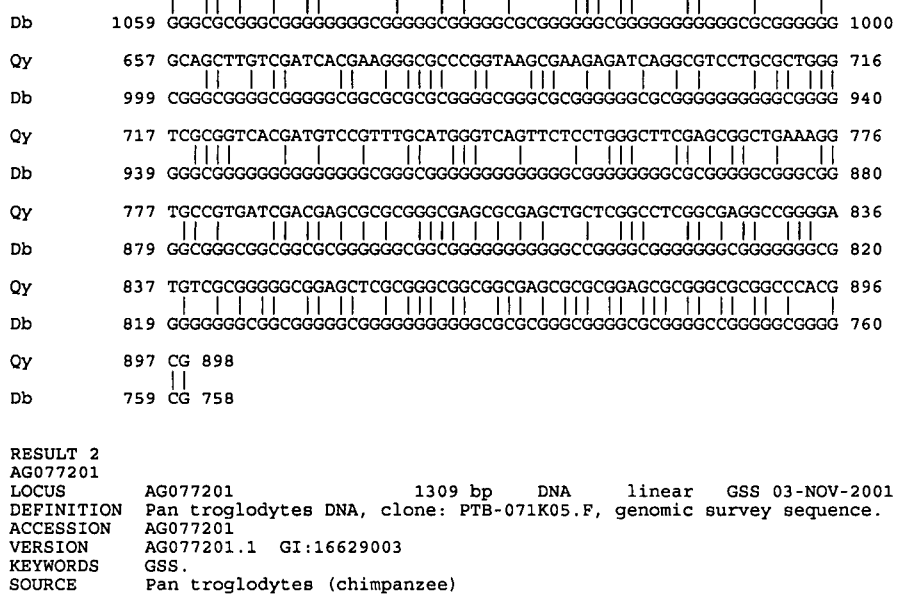
ALIGNMENTS

RESULT 1
CL078342/c
LOCUS CL078342 1674 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-149D22_Sp6.1 CH216 Xenopus tropicalis genomic clone
CH216-149D22, genomic survey sequence.
ACCESSION CL078342
VERSION CL078342.1 GI:40534255
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1674)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 300
High quality sequence stop: 387.
FEATURES Location/Qualifiers
source 1. .1674
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-149D22"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"
ORIGIN

Query Match 10.8%; Score 97.2; DB 10; Length 1674;
Best Local Similarity 44.8%; Pred. No. 2e-08;
Matches 404; Conservative 0; Mismatches 494; Indels 4; Gaps 1;



RESULT 2
AG077201
LOCUS AG077201 1309 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-071K05.F, genomic survey sequence.
ACCESSION AG077201
VERSION AG077201.1 GI:16629003
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1309)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .1309
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-071K05.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN



```
Qy        494 TATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACATC 553
              |  || |  | | |||     |  ||  || | || ||   |||   |       |
Db        705 ACCCGGCCCGGGCGGGGCGCGGGGGGAGGCGGGGGCGCGGCGGGACGCGCGGGCCGGGGC 764

Qy        554 GGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCCCGGAAC 613
              ||   |      || |||          ||          ||    ||| || || | | |
Db        765 GGGCGGCGCGCGGCGCGCGACGCGGCGGGCGCGAGCAGCACNCGCGCGCGCGCGCCCGCGG 824

Qy        614 AGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCGATCACG 673
              |     | ||  |  ||   | | || |   |  |  |  | || |          .| ||
Db        825 GGGCGGGCGGCGNCGCCCGCCGCGNCCGCCCCCCCGCGCGCGGGCGGACGCGGCGGCGCG 884

Qy        674 AAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACGATGTCC 733
              | |||    |   |||       | ||   |   ||  ||| |||||   |   |  |
Db        885 CGCGCCGCGGCGGCCGCGGGCGCGCCCGGGCCCGCCCGGCGGGCCGCGGGGCCCGGGCGC 944

Qy        734 GTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCGACGAGC 793
              |    | |||   | | |   | |   | ||||| |    | |  ||    || || ||
Db        945 GGCCCCGCGGGCGGGCCCGCGCCGCCGGGGCGCGGCCGCGGCGCGACGCCGGCGCCGGGC 1004

Qy        794 GCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAGC 853
              |    ||| ||| |  | |    | |   ||  | | || |  | |||||||| |||| 
Db       1005 CCCGCCGCGGGCGGGGCCCGGCGCGGCGGCGCCCGCCGGGCGGCGGCGCGGGGGCCGAGC 1064

Qy        854 TCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGTC 900
              ||||  ||      | |||||||  ||||||    || |  |||  |
Db       1065 GCGCGCCCGCGCCGGCGCGCGCGCCGCGCGGCGAAGGGCGGCGCNCC 1111
```

RESULT 3
AG363333/c

```
LOCUS       AG363333               2332 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-165A22.TJ, genomic survey
            sequence.
ACCESSION   AG363333
VERSION     AG363333.1  GI:47974538
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 2332)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
```

```
FEATURES             Location/Qualifiers
     source          1. .2332
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-165A22.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

Query Match 10.5%; Score 94.4; DB 10; Length 2332;
Best Local Similarity 44.8%; Pred. No. 6.6e-08;
Matches 404; Conservative 0; Mismatches 496; Indels 2; Gaps 1;

```
Qy          2 GCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGC 61
              ||| || |  ||   | |||   | || || |  | ||| | ||  || | ||
Db       1886 GCGCCGGGCGGGCGCGCGCGGGGCCGCCGCGCGCGCCGGGCGAGCCGGCGGGCGCGCGCG 1827

Qy         62 GCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGC 121
                | || || | |   |  | ||| ||||  || |    | |       ||| || | 
Db       1826 CGGCGCGGCGGGGAGGGGCCCGGCGGCGGGGGCGCGGGCCCGGGCGCGCGCGCCGGCC 1767

Qy        122 GCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAG 181
              ||  ||| ||    |   |  | || |||||  | || ||| |   |      |
Db       1766 GCGGGGCCGACCCGGGGGCCCGGGGCGCGGTCGCGCCCCGGGCGGCGGCGGGGGGCCGGG 1707

Qy        182 ATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGC 241
              ||||   || | |     |||| ||          || ||| |   | |||||  | ||
Db       1706 CGGCCGCCGCGGGGGCGCGCGCGGCGCCGCCGGCGGCCGCGCCGGCGGGCGCGGGGGGGC 1647

Qy        242 ACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGA 301
              ||||  |||       |   ||||   | |  |    ||   |   ||| || | 
Db       1646 GCGGCGGCGCGCGGCCGGGCGCCCGCGCGCCGGGGCGCGTCCCCCGCGCCGGCCGCGGGC 1587

Qy        302 TCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGC 361
              |||    |  |    |  |||  |   |  |||  || |  | | |||||||  |
Db       1586 GGCCGGCCGGCCGGGGCGGCCGCGCGGGCCCCCCGGGCGGGGCCCGGCGCGGCGGGCCGG 1527

Qy        362 GGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGC 421
              |||  ||||  |   ||          |   ||  |  |   |  |  ||| |||| |||
Db       1526 GGCGGGCCGCCGGCCGCCGCCCCGCGGCGGCCGCACCGGCGGCCGGGGGCCCCGCGCCGC 1467

Qy        422 GCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGG 481
              |||||||||||||  ||   ||   | |||  | || ||  | |  |||  |     |
Db       1466 GCGCCCGCGGCCGCGGCCGCGCGGCGGGCGGCGGGCCGGCGGGGGCCGGGGGCCGCCGCC 1407

Qy        482 CAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGC 541
              |    | |||||    ||  | |  |  | | |  |  |||     |   |  | ||| |
Db       1406 CGCGCCCCGGCGCCCCCGGCGGCGCCGGCCGGGGGCGCGCGCGCCCCGGGGGGGGCGGCC 1347

Qy        542 AGCGAGAACATCGGATAGGCGCGGGAG--CGGTCGATCTCGACGAGCCAGAGATAGCGCA 599
              |  |      |  ||   |||     |||    |  ||    |  |   | ||
Db       1346 GGGCGGCCGGCGCGCGCGGGCGGGGGCGCCGGCGCGCCCCCGCGCCGGCGCGGCCGGGCC 1287

Qy        600 CGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCA 659
              || ||   |||    | ||  | ||  ||  |      |    || |  |     | | 
Db       1286 CGCGCCGGCGGCGGGGGGCGGCGGCGGCCGGCGCACCCGGGGGGGGCGGGGGGGGGCCCG 1227

Qy        660 GCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCG 719
              |     | |  |||||  |||  | | |  |    ||  | | |||  |   |
Db       1226 CCCCCCCCCCCCCCTCGGGCGGGCGGGCGGGGGGGCGCGGGCGCCGGCGGGGCGGGCCCC 1167

Qy        720 CGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGC 779
              |||   |   ||   | |  |       | |  || | ||  |   | |
Db       1166 CGGCGGGGCGCCGCGCCCCGCCGCGGGACGCAGGGGGCGGGGCGGGGGGGGGGGTGCCCC 1107

Qy        780 CGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGT 839
              |     |  |  | ||||||  |  ||| || |  |||     | | ||||| |
Db       1106 CCCCCCCCCCCCCCCCCGCGGGGGGCGGCGGGCCGGCCGGGGCGCGGGCGGCCGCGCGGCG 1047

Qy        840 CGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGCGT 899
```

```
Db       1046 CCACGCCGGCGAGGGCGCGCGGGCCGGGCGGCGCGCCGCGCCGGCGCGCGGCGGCCGCGT 987

Qy        900 CG 901
              ||
Db        986 CG 985


RESULT 4
AG435689/c
LOCUS       AG435689               1713 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-313O06.TJ, genomic survey
            sequence.
ACCESSION   AG435689
VERSION     AG435689.1  GI:48078752
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1713)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1713
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-313O06.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             10.4%;  Score 93.8;  DB 10;  Length 1713;
  Best Local Similarity   45.4%;  Pred. No. 8.9e-08;
  Matches  381;  Conservative    0;  Mismatches  450;  Indels    8;  Gaps    2;

Qy          58 CGGCGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGAC 117
               |||||||   ||| |  | ||||| |||| || |        | |||| | |    |  |
Db        1637 CGGCGCGCCGCGCGGCCGGCGCGCCGGCTCCGCGCGCCTCGCGGGCGTGCTCCGCCCCGC 1578

Qy         118 GAGCGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGC------CGGGGAGCTCGCG 171
               | || | |||| |   ||   |||| |  || || || | |      | ||  ||   ||

Db        1577 GCGCCCGCGGCCGCCCGCCGGCCGGCCCCCGCGGCCGCGCCGNGCCGCCGGCGGCCGCCG 1518

Qy         172 CACGTAGAAGATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGAC 231
               | ||  |  |  ||    || |  |  |   |||    | | | |  |||      |  |
Db        1517 CGCGGCGCGGCGGCGCGCCCCCCGGGCGGCCCGGCCGCCCGGCGCGGCCCCGGCNCCGC 1458

Qy         232 GCGGACGAGCACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGA 291
               | | ||    |  |  ||||| |       | ||| |    || |||     | | |
Db        1457 NCCGCCGCCGGCCCCGCCGCCGGCCCGCGCCGCCGGCCGCGCGCCCCCGCCGCCGCCGCC 1398

Qy         292 CGCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACAC 351
               |||| ||||  | | |  | | || ||  |||  | | | ||||||  ||| || |
Db        1397 GGCCGCGCGCGCCCCGCCCGCCCCCCCCGCCCCGGGCGCCGGCGCCGCGCGCCGCCG 1338

Qy         352 CGCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGC 411
               |  | ||||    |  |      || ||| |  |||| ||  |    ||  ||  | ||
Db        1337 CCGGCCCGGCCCGCCCGCCCCCGGCCGCCGCGCGCGCCCCGGCCGCCGCGCCGCGCCGN 1278

Qy         412 CTCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAA 471
               | |||    |  ||| |  ||||         ||      || ||  |  |  |||||||
Db        1277 CCCGCCCGCCCNGCCGGGCGCCGCGCGCCCGCCGCCGCCGCGGCCCGCGCGGCGCCCGCC 1218

Qy         472 GAAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTC 531
               |   || | || |  | | ||   | ||  |   |  | | |  || || |
Db        1217 GCGCCGGCCGCCG--CCGGCCGNCGCCCGGCCCGCGCCGCCCCCGCCGCGCCCCGGGCCG 1160

Qy         532 GTCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAG 591
               |||  |  |     |   |  |||||    || | |     |  || ||    |
Db        1159 CGGCGCGCCCCNCCCCCGCCGGGCCGCGGCGCCCCCGCCGCCCCCGGCCGCGCGCGCCGG 1100

Qy         592 ATAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTT 651
               ||||  ||  |  ||| |  |  | | |  || ||  || |  ||||   | |
Db        1099 CCCGCGCCCCGCGGCCCCCGCGCGCNGGCCCGCCCGCCCGCCCGCCGGCCGGCGCCCCCCC 1040

Qy         652 CATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCG 711
               |     | ||    |  | ||   |||||       ||          | || |   ||
Db        1039 CGCCGCCCGCGCCCCNGCCGCGCGCCGCGCCGCCCGCCCGCCCCCGCGCGCCGCCGCCCG 980

Qy         712 CTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTG 771
               |  || |||      ||  |  |      |  || |      |  ||||  || || || |
Db         979 CGCGGCCGCCCCGCCGCCGCGCCGCGCCGGCGGCCGCCGGCCGGGGGCCCCGCGCCGCGG 920

Qy         772 AAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCC 831
               || ||||    ||  |  || || | | | |  |    ||     | | ||
Db         919 GCCGGCGCCGGCCCCGCGCGCCCCGCCGCGCCGGGGGGCGGGGGGGGGGGGNNGGGGGGGG 860

Qy         832 GGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGG 890
               |  |  | |||| ||  |  | ||| || || | | | |  | | | ||| | ||
Db         859 GNNGGGGGNGGGGGGNNGGGCCGGGGGGGGGGGGGGGGGGGGGNNGNGGGGGGGGGGGG 801


RESULT 5
AG435170
LOCUS       AG435170               1822 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-313C14.TJ, genomic survey
            sequence.
ACCESSION   AG435170
VERSION     AG435170.1  GI:48078233
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
```

```
REFERENCE   2  (bases 1 to 1822)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1822
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-313C14.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match            10.4%;  Score 93.8;  DB 10;  Length 1822;
  Best Local Similarity   44.7%;  Pred. No. 8.8e-08;
  Matches  398;  Conservative    0;  Mismatches  485;  Indels    7;  Gaps    2;

Qy          2 GCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGC 61
              | ||||||     |   || || |  ||     |   | | |       || ||  |   ||
Db        875 GGGGGGGCGGGGGGCGGGGGGGCGGGCGGGCGGGGGGGGGCNCGGNCGGGGGGGGGGGG 934

Qy         62 GCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGC 121
              | ||  ||  | |   | | |||    || | |  | || |||  || | |||| || |
Db        935 GGGGCGCGGGGNGCGGGGGGGGGGGNGGGCGGGGGGGGGGGCGGGCGGGGGGCGGCGCGG 994

Qy        122 GCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAG 181
              | | | | |  ||       ||  | || || || ||| ||||  |    |       |  |  |
Db        995 GGCGGCCGGGCGGGCCGGGGGGGGGCGGGGGCGCGGCGGGGGGGCGGGCGGGGGGGGGG 1054

Qy        182 ATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGC 241
                || | || | | |      | ||  || || |  |   ||  |  ||  | ||  | ||
Db       1055 GGGCGGGGCGGGGGGCGCGGGGGGGNCGGGGGCGGCGGGCGGCGGGCGGGGGGGGGGGC 1114

Qy        242 ACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGA 301
              |||  ||  |  | ||    | ||   || |              |  || || | || |
Db       1115 GGGGCGGCGGGGGGGGCGGCGCCGGCGGCGCGGGGGGGGGCGGCGGGCGGCGGCGCGGGG 1174

Qy        302 TCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGC 361
                ||  |  |||| |      |     ||  ||  |   ||      | ||| ||| |
Db       1175 GGCGGGGGCGGGCGCGGGCCCGGGGGGNCGCGCGGGGGNCCGGGGGGCGCCGGGGGGGCG 1234

Qy        362 GGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGC 421
              || ||  | ||| |  |   || | |  | ||    | |  |  |       || |
Db       1235 CCGGCGGCGGGGCGGGGGGGGCCGGCCGGCGGGGGCGGNGCGCGGGGGGGNGGCGGGGGG 1294

Qy        422 GCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGG 481
              | |  ||||| |||    |  || |||||||  |  || ||  |  | |       |  ||
Db       1295 GGGGGCGCGGGCGGGCGGGGGCGCGGCGCGGGCGGCGGGGGGGCCGGGG------GGGGG 1348

Qy        482 CAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGC 541
              |  |  ||||    || ||  |  | ||  |     |  | | ||     |      |||
Db       1349 GCGGCCGGGGCGGCGGCGGGCNGGGCGGGGCGGGCGGGCGGGCGGGGGCGGGGNGGGCGG 1408

Qy        542 AGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACG 601
              |||  |      ||   |||| ||| |||  ||    |   |     |     |||  |
Db       1409 GGCGCGGGGGGGGGGGGGGCGGGGGGGCGCGCGGGGGCNCGGGCGGGGGCGGCGCGGGGG 1468

Qy        602 AGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGC 661
              |   | |    | ||  | ||       | |||  |      || |     |  |
Db       1469 GGGGGCGCGGCGGGGGCGGCGGCGGGGGGCGGGGGGCGGNGGGNCGGCGGCGGGGCGGCG 1528

Qy        662 TTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCG 721
              |  |  |  |  |||||| |      |    | ||  | ||  |  |  |  | | ||| |
Db       1529 GGGGNGCGCGGGCGGGGCGCGCCCGCGGGCGCGGGGGCGGNGGGGCGCGGGGGCGGCGGG 1588

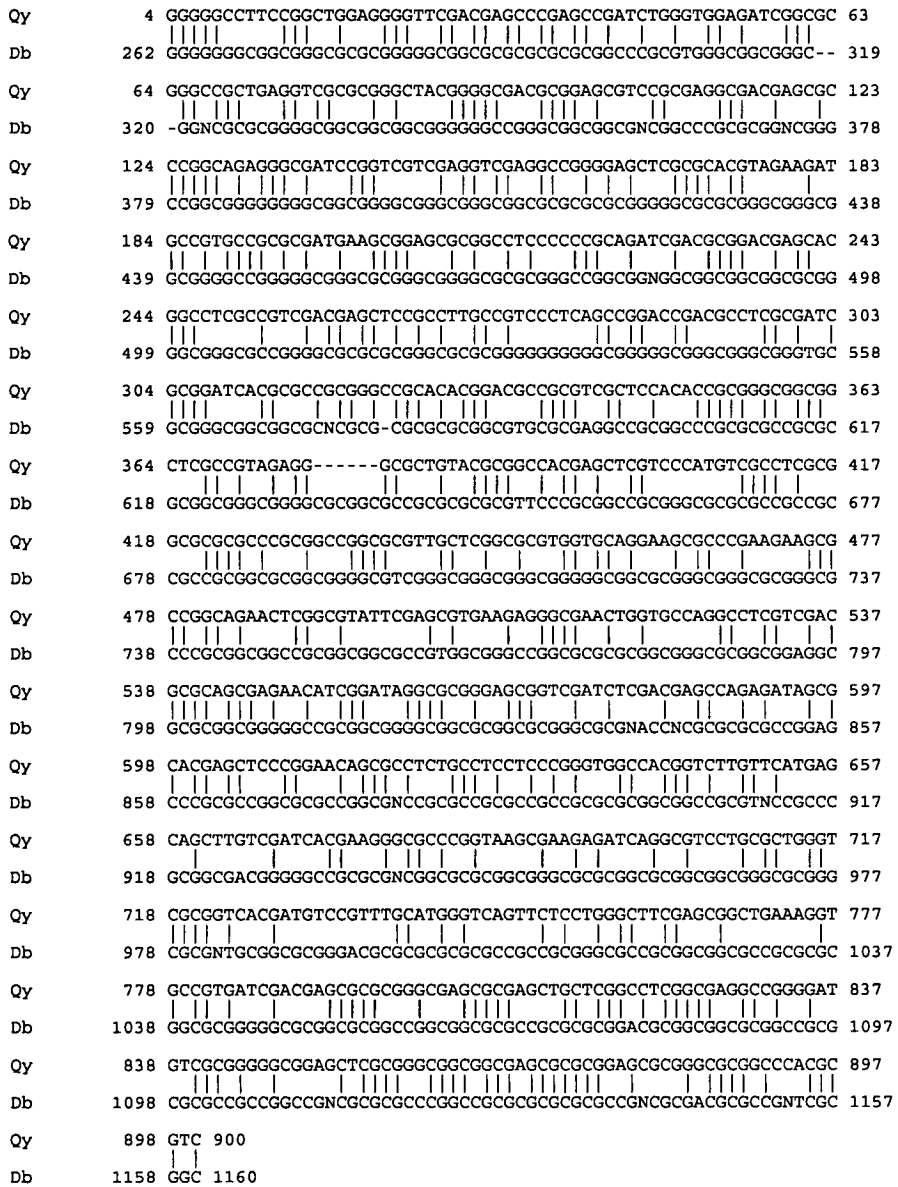
Qy        722 GTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCG 781
              | | ||  |      ||  ||    |  | |  ||     |||   |   |  ||
Db       1589 GCCGCGGGGGGNCGCGGCGCGGCCGCGNGCGCGGCGGGCGGCGGCGCGCGGGGCGCCGCG 1648

Qy        782 TGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCG 841
              |  || |  | ||| | ||| ||| | |  ||   ||  |||  |  |  || |  | ||
Db       1649 CGGGCGGCCGGNGCGNGCGCG-GCGGGGGGNGCGGCGCGGCGGGGGCGGGGGCGCGGCCG 1707

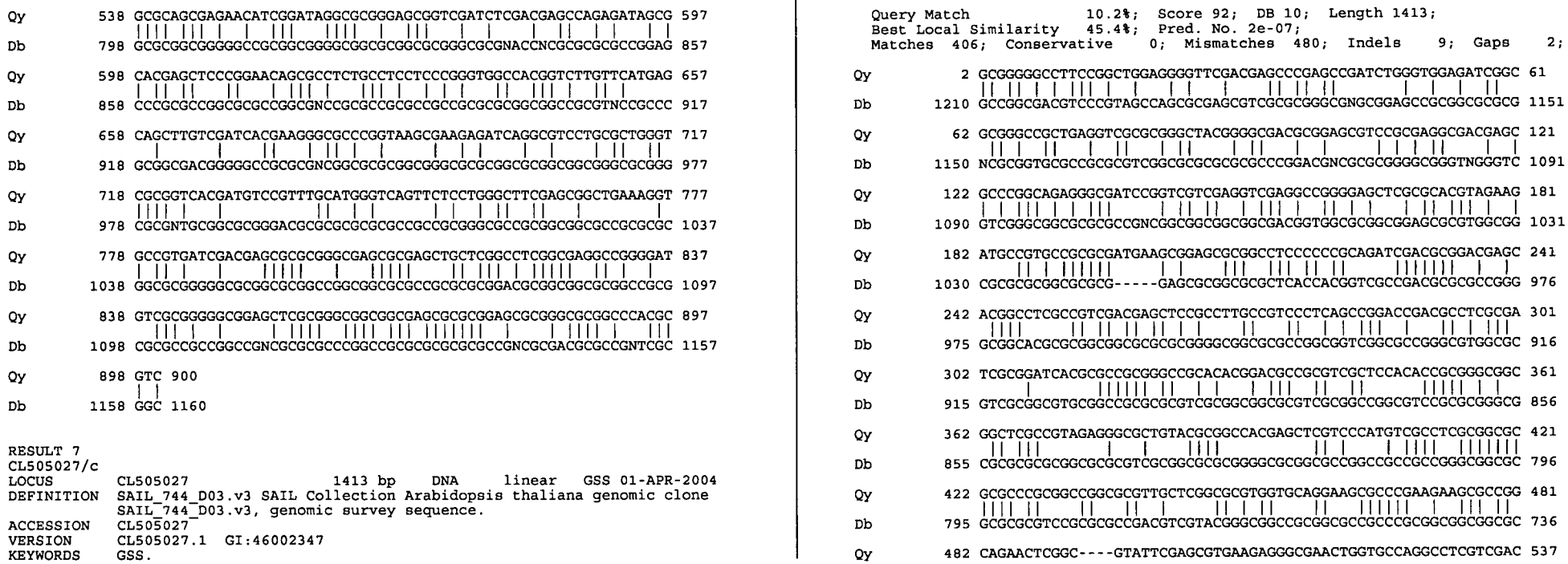
Qy        842 CGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGC 891
              ||| | || |  |||||| ||| ||| ||| |||| ||| | | |||||
Db       1708 GGGGCGGGGGGGGCGCGGGGGGCCGCGGGCGGGCGGCGCGNGCGNGCGGC 1757

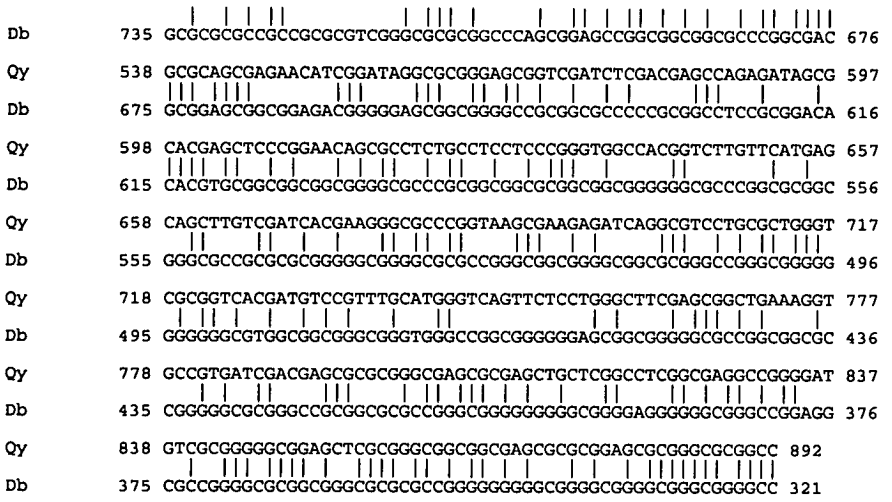

RESULT 6
CL470317
LOCUS       CL470317               1280 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_140_A11.v1 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_140_A11.v1, genomic survey sequence.
ACCESSION   CL470317
VERSION     CL470317.1  GI:45936028
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1280)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS806754; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1280
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_140_A11.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match            10.4%;  Score 93.4;  DB 10;  Length 1280;
  Best Local Similarity   45.7%;  Pred. No. 1.1e-07;
  Matches  413;  Conservative    0;  Mismatches  480;  Indels   10;  Gaps    3;
```



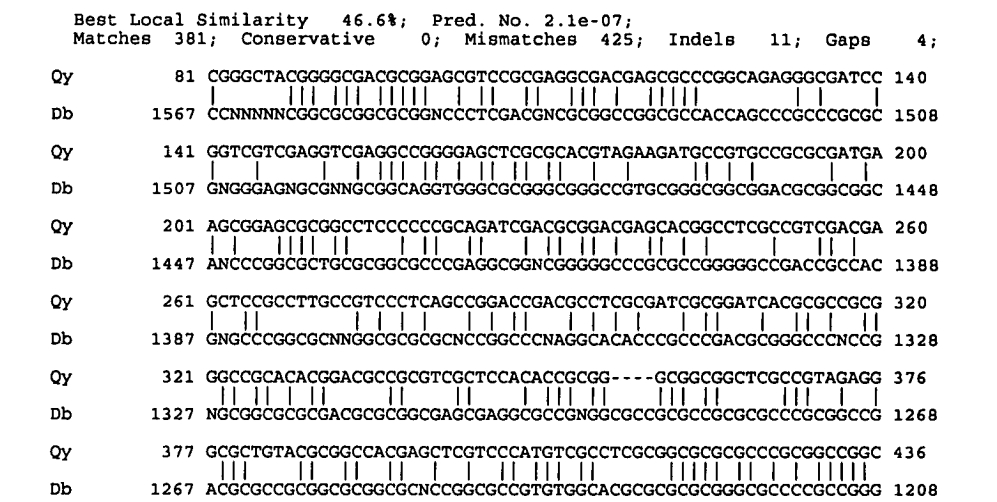
RESULT 7
CL505027/c
LOCUS CL505027 1413 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL_744_D03.v3 SAIL Collection Arabidopsis thaliana genomic clone SAIL_744_D03.v3, genomic survey sequence.
ACCESSION CL505027
VERSION CL505027.1 GI:46002347
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1413)
AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B., Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
TITLE A high-throughput Arabidopsis reverse genetics system
JOURNAL Plant Cell 14 (12), 2985-2994 (2002)
PUBMED 12468722
COMMENT Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number CS833289; T-DNA left border flanking sequences of Syngenta Arabidopsis Insertion Library (SAIL) lines are available through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not single contiguous sequences.
Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1. .1413
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_744_D03.v3"
/clone_lib="SAIL Collection"
/note="T-DNA left border sequences were isolated using a modified TAIL-PCR strategy"
ORIGIN





RESULT 8
CL509291/c
LOCUS CL509291 1694 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL_811_C05.v3 SAIL Collection Arabidopsis thaliana genomic clone SAIL_811_C05.v3, genomic survey sequence.
ACCESSION CL509291
VERSION CL509291.1 GI:46006597
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1694)
AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B., Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
TITLE A high-throughput Arabidopsis reverse genetics system
JOURNAL Plant Cell 14 (12), 2985-2994 (2002)
PUBMED 12468722
COMMENT Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number CS836232; T-DNA left border flanking sequences of Syngenta Arabidopsis Insertion Library (SAIL) lines are available through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not single contiguous sequences.
Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1. .1694
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_811_C05.v3"
/clone_lib="SAIL Collection"
/note="T-DNA left border sequences were isolated using a modified TAIL-PCR strategy"
ORIGIN

Query Match 10.2%; Score 91.8; DB 10; Length 1694;



RESULT 9
AG171124
LOCUS AG171124 1798 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey sequence.
ACCESSION AG171124
VERSION AG171124.1 GI:16700802
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.

```
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match            10.2%;  Score 91.6;  DB 10;  Length 1798;
  Best Local Similarity   43.0%;  Pred. No. 2.3e-07;
  Matches  388;  Conservative    0;  Mismatches  502;  Indels   12;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              ||    ||     ||||  | | ||     || | |      |    | | |     ||
Db        596 GGGNNNGGGGNGGCGGCGCGCGCGGGCGCGCGCGGCGCGCGGGGNGCGCGGGCGCCGCGC 655

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
                | || |  | || |||| | ||    ||    |||||      |||| |    || |
Db        656 GCGGCGCGGGCGCGGCCGCGGGCGCCCGCGGCGCGCGCGGNCGNGGCGCGCGCGCGCGCG 715

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              ||| ||   | | |||  | |  || || |  || |  ||  | || ||||| ||  |
Db        716 CGCGCGCGCGCGCGCGCGGCGCGGCGCGCGCGCGCGCGNCGCGCNCGCGCGCGCCGGN 775

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
                |   || ||||       | | |||| ||| |  | |   |  || ||||  || |
Db        776 CGCCCGCCGCGGCGCCCGCNCNCCGCGCGCCGCCGCGGCGCNNNGCGCGCCGCGCNCGCG 835

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCG--------GACCGAC 292
              |   ||  ||||  ||  | || |   |  ||    || |   ||        | |   |
Db        836 CCGCGCGNCGCCCCCGCGGCGCGCNCGCCCGCGNCGCCGCGCGCGCNCGCCGCGCCGCGC 895

Qy        293 GCCTCGCGATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACC 352
              ||| ||||  ||||     |||| ||| |  | | |   | |  | || ||| |  | ||
Db        896 GCCNCGCGCCCGCGCGNNCCGCGGCGCCGCGCCCGCCNCGCCCGCNCGGCGCNCNNCGCC 955

Qy        353 GCGGGCGGCGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCC 412
              ||     ||  | |||||  | ||     |    | ||      ||     ||       |
Db        956 GCCCCGCGCCNCNCGCCGCCGCGGCGCNCGCCGCCCGCGCGCCGCCGCCGCCCNCGGCGC 1015

Qy        413 TCGCGGCGCGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAG 472
              |  |  || ||||||||  |          |||||| |     |   |||| ||
Db       1016 GCCNGCNGCNCGCCCGCGCGCCNNNNNNGCGCCGGCGCGCGCGCGCGNGCGCGCGCGCGC 1075

Qy        473 AAGCGCCGGCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCG 532
                |||   |    | |  ||                | | |||  |  | |||    | ||
Db       1076 NNNCGCGCCCGCGNCNCNCCGCGNCGCCCGCCCGCGCGCGCGNCCGCGGGCCGCCNCNCG 1135

Qy        533 TCGACGCGCAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGA 592
              |  |          |  |  |||    |||| | |||  ||    || || ||    |
Db       1136 CCCGCCGCGCNNCCGCNCNGCGGCGGCGCGCGCGCGCGCGCG---NCGGCGCGCGCCGGC 1192

Qy        593 TAGCGCACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGC-CACGGTCTTGTT 651
              || | ||    | |    | ||||| | ||  |   | | |  || | ||| |  |
Db       1193 GCGCNCGCGGCNNCGCCCGCCGGCGCCGCCGCGCCGCGCGGCGNCGCGCGCGGGCGCGCN 1252

Qy        652 CATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCG 711
              |  |  | ||   ||  | |        ||||          |  |   ||    |||
Db       1253 CGCGCCCGGCNGNGCGCGCNCCCGCNCGCGCCGGCCGCGCCCCCGCGCGCNCGCGGCGCG 1312

Qy        712 CTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTG 771
              |  |  ||||| |      |  ||        |||       | |  | |   || |  || |
Db       1313 CGCGCGCGCGGGCCGCGCGCGCGGCGCGGGGCGCCGGCGCGCGCGGCGCGCGGCGCCG 1372

Qy        772 AAAGGTGCCGTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCC 831
                || |||| |        || |  | ||         |  | ||  |    || |||
Db       1373 CGCGGCGCCGCGGGGNCGCCGCNCCGGCGCCGCGCGCGCCNCCGCGCNCGGNCCGCGGCG 1432

Qy        832 GGGGATGTCGCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGC 891
              |  |    ||     |      |||| ||  | |||||||  |       ||| ||
Db       1433 GNCGNCCGGCCGCCNCNCGGCNCGCGCGGCGCCGCGGCGCGCGCGCCGGCGCCNCGCCGC 1492

Qy        892 CC 893
                |
Db       1493 NC 1494


RESULT 10
BQ678719/c
LOCUS       BQ678719               1288 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8199257 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6261563
            5', mRNA sequence.
ACCESSION   BQ678719
VERSION     BQ678719.1  GI:21791398
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1288)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2422  row: h  column: 12
            High quality sequence stop: 171.
FEATURES             Location/Qualifiers
     source          1. .1288
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6261563"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
```

Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 10.0%; Score 90.2; DB 5; Length 1288;
Best Local Similarity 42.0%; Pred. No. 4.3e-07;
Matches 368; Conservative 0; Mismatches 507; Indels 2; Gaps 1;

```
Qy         16 GCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGCCGCTGAGG 75
Db       1137 GCGNGCGCGGCCGGCGGGGNCGGGCCCCGCGCGGCGGNNGGGNNCGCGCGGCCGCGCCGC 1078

Qy         76 TCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGGCAGAGGGC 135
Db       1077 NCGCGNCGNCGCGCGCGGGCGCGNCCGCGCGCGCGGCGCNCGCCGGGGCGCCGCGGCGGC 1018

Qy        136 GATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCGTGCCGCGC 195
Db       1017 GGGCGGGCGGCGCCCGNCGCGGGCGGGGCGNNGCGCGGCG--GCGGCGNNCGGGCGGGGG 960

Qy        196 GATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCCTCGCCGTC 255
Db        959 GCCGCGGCGGCCGGGCGGCGGCCGGCCCCNCCNCCGGGGCGCGGCCGCGGNGCGCGCGGC 900

Qy        256 GACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGGATCACGCG 315
Db        899 GGCGNGGGNGGGCCNGCGCGCCCCCGCGGGGGGGGGCGGCGCGGGNGGGCGGGCCGCNGG 840

Qy        316 CCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCGCCGTAGAG 375
Db        839 CCGCGCGGCCGGCGGCCGCNCGGCCGGGCGCGGCNCNCCGCNGGGGCGCNCGGCGNCGGG 780

Qy        376 GGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCCGCGGCCGG 435
Db        779 CGCGNGNCGGGCGCGCNCGGGCGCGCGCGGNNCCNNGGGCGNGNGGGCCGGNNGNGGCGG 720

Qy        436 CGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCAGAACTCGGCGTA 495
Db        719 CGGGCGGGGGCGCNCCCCGCGGGGCGGGCCCNGGGGCGGCCCGCGGGGGGCCCCCGCCGG 660

Qy        496 TTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAGCGAGAACATCGG 555
Db        659 GGGGCGGGGCGGNCCCGGCGGGGCNNGCGGGGGGGCGGGGGCGGGGGCGCGGNCCGCCCC 600

Qy        556 ATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAGCTCCCGGAACAG 615
Db        599 GGCNGCCGGGGGCCGCCCGNGGGCGGGGCGCGGGGNCGGGCGCGCGNCCCGCCGGGCGCG 540

Qy        616 CGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTTGTCGATCACGAA 675
Db        539 CGCGGGCGGGGGGCCCCGGGGCGGGCNNCGGGGGGNGGCGCGCCGGCGGCCCGGGGCGGG 480

Qy        676 GGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGTCACGATGTCCGT 735
Db        479 GGGGGGGGGGGGGGCGGGGGGCNCGGGCGCGCGCCGCGGGNGGGGGGCGCCGGGGCCCCC 420

Qy        736 TTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTGATCGACGAGCGC 795
Db        419 GCCGCGGCGGGGCGGCNGGGCGGCCCGCGGCCCGGGGGNGGGGGGGGGGGGGGACCGGGG 360

Qy        796 GCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCGGGGGCGGAGCTC 855
Db        359 GCGGNGGCGGGCCCGGCGGGGGCCCGGGGCGGGGGGGCGCGGGGNGGGGGGNGGGGGGC 300

Qy        856 GCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCC 892
Db        299 GGGGGGGGGGGGGANGGGGGGGGGGGGGGGNGGGGC 263
```

RESULT 11
AG441877
LOCUS AG441877 1462 bp DNA linear GSS 22-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-323H09.TJ, genomic survey sequence.
ACCESSION AG441877
VERSION AG441877.1 GI:48084940
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1462)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1462
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-323H09.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 9.9%; Score 89.2; DB 10; Length 1462;
Best Local Similarity 46.7%; Pred. No. 6.6e-07;
Matches 423; Conservative 0; Mismatches 458; Indels 24; Gaps 5;

```
Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
Db        358 GGGNGGGNNGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGNGGNGGGGGGGGGNGGGGGG 417

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
Db        418 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGNGGGGGGGGGGGGGGGGGGGGG 477

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
Db        478 GNGGNGNGGGNNGGGGGGGGGGCCCCCCCGGNGNGGGNNGGGGGGCCCCCGCGCGGGGGG 537

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
Db        538 GGGGGGGGGGCGCGGCCGGGGGCCCGCGCGGCGCGCCGCCGCCCGCCCGNCGCCCGGGCG 597

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
```

```
Db        598 GGCGCGGGGGCGCGCCCCCGCGGGCGCCCCGCCGGCCCGGCGCCGCGCCCCGGCGCGGCC 657

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360

Db        658 CCCGGGGGCCCCCGCGCCGCCGCGGGCGGGGCCCCGCGGCGGCGCGGCGNCGGCGCGGGCG 717

Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420

Db        718 GGGCGCGGCGCCGCCCGCGCCGCCGCGGGGGCCGCCGGGCCC-------GCGGCCGCGCGCGG 770

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480

Db        771 CGGCCGGCCGGCGGGGGCGGCGGGGGGCCGGCGCCGGGGGCCGCGCGGGCCCGCGCGGCC 830

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540

Db        831 GCCGGCGCCCGCGGCGCCGGCCGCGCGGCCGGCGGGCGGGGCGGCGGGCGGGCCCGCCGG 890

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCG--- 597

Db        891 CCGGGCGGGGGGCGCCCCGGGGGCGGCGGCCCGGCGCCCGGCCCGCCGGGGCCGGGGGGG 950

Qy        598 --------CACGAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTG 649

Db        951 CGGGGCGCCCGGCGCGCGCGGGGCGCCGCCGCGGCCGCCCGCCGGGGCGCCGCGCGCGCG 1010

Qy        650 TTCATGAGCAGCTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTG 709

Db       1011 GCCGCGGGCGGGGGCGCGCCCGGNCGGGGGCGCGG----GCGGCCGGGGCCCGNCGCGCG 1066

Qy        710 CGCTGGGTCGCGGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGC 769

Db       1067 CGCCGCGGCCGCGCCCGGCGGCGCGGCGCGCGGCGGGCCGCGGCCGGGGCGGCGNGGGCC 1126

Qy        770 TGAAAGGTGCCGTGATCGACGAGCGCGCG-GGCGAGCGCGAGCTGCTCGGCCTCGGCGAG 828

Db       1127 GGGGCCCCGCCCGGCGCGGCGCGCGCGCGGCGCGCGCGCCGGCCGGCCGGCGGGGC 1186

Qy        829 GCCGGGGATGTCG-CGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCG 887

Db       1187 GCGGGGGCCCCGGCCCCCCGCGGCGCGCCCGCGCCGGGCCGGGCGCGGCCCGGGCCGCGG 1246

Qy        888 CGGCC 892

Db       1247 CGGNC 1251
```

```
RESULT 12
BG809984
LOCUS       BG809984               1516 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  mgct002xd11f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct002xd11f 5', mRNA sequence.
ACCESSION   BG809984
VERSION     BG809984.1  GI:14180964
KEYWORDS    EST.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES             Location/Qualifiers
     source          1. .1516
                     /organism="Magnaporthe grisea"
                     /mol_type="mRNA"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct002xd11f"
                     /dev_stage="Germinated conidia on apprressorium-inductive
                     surface"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed from conidia germinated for 5-8 hr
                     on an inductive surface. The library contains over 55,000
                     clones with average insert size of 1.5 kbp."
ORIGIN
```

```
  Query Match             9.8%;  Score 88.6;  DB 2;  Length 1516;
  Best Local Similarity   43.3%;  Pred. No. 8.5e-07;
  Matches  375;  Conservative    0;  Mismatches  483;  Indels    9;  Gaps    2;

Qy          2 GCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGC 61

Db        355 GCGCGCGCGGGGCGCGCGCCGCCGCCGGNCCCGNCGCNCCCNCGGCGCGGGCCGCGCGCG 414

Qy         62 GCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGC 121

Db        415 GCNGCGGCCGGCGCGCGCCGCCGCCGNCCGCGCGGCGNGCGGCGCGCGCGCGCGCGCGC 474

Qy        122 GCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAG 181

Db        475 GCCCGCGGGCGCCGGNGNGNNGGGGNGCGGCCGCGGGGGGGGGCGGNCNCCNGCNCGCC 534

Qy        182 ATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGC 241

Db        535 GCGCGGCGCCGCCGCGGGCNGCCGCGCGCGGCGCGCGCNGCGGGGCGGGCNCGCG---GC 591

Qy        242 ACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGA 301

Db        592 CGGGCGCCGCGGGGGCCGCGCGGCGCCCGGCGCCCGCCCNGCGCCGCCCGCCNCGCGC 651

Qy        302 TCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGC 361

Db        652 CCGGCGCCCGCGCCGCGCCCGCCGGCGCGGGCCGGCGNGNGCGCGGCGCGGCCGGCGGCG 711

Qy        362 GGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGC 421

Db        712 GGCGCGCGGCGGGGCGGGGCGCCGNCGCGGNNGGGGGCCCCNCGCGCGGCCGGCGGNGCC 771

Qy        422 GCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGG 481

Db        772 GGGCGCGGCGGCNGGCCGCCGCGGCGNGCGCGCCG------GCGCGCGCCGCCGCGNNGC 825

Qy        482 CAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGC 541

Db        826 GCGCCAGGCGCGCGCGCGGGGCGGGGGGCGGNCGGCGCGCGCCGCGCGGGCNGGGCGCGG 885

Qy        542 AGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACG 601

Db        886 GGCGCGCGCGCCGCGGCGCGGCCGCGGCNCGCNGCCGCGGCCGCCCCGCGCCGGGCCGCC 945

Qy        602 AGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGC 661

Db        946 CCGGCCCCGCGCGCCCCCCCGCCGGCGGCGCGGGCGGCCGCGGCGCCGCGCCGCGCGGCC 1005

Qy        662 TTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCG 721

Db       1006 CGCGCCNCGCGCCGCGGCGAGCCGCNCCGGACGCCCGCGGCCCGCCGCGCGCGCCCGCGGC 1065

Qy        722 GTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCG 781

Db       1066 CGCCCGGCGCGCGGCGCCGNGNNGGCGCGCCCCGCGNGCGCGGGNGCCCGCCGCGCGCNC 1125

Qy        782 TGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCG 841
```

```
Db       1126 GCNGCCGCGCGCCCNCGNGCGNGCCCGCGCGCGNNNCNCNCGCCGCCGCCGCACGCGCGC 1185

Qy        842 CGGGGGCGGAGCTCGCGGGCGGCGGCG 868
Db       1186 GCCCGNCGCCGCGCGCGCGGAGCCGCG 1212


RESULT 13
AG448267
LOCUS       AG448267               1523 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-332C15.TJ, genomic survey
            sequence.
ACCESSION   AG448267
VERSION     AG448267.1  GI:48091330
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1523)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1523
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-332C15.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             9.8%;  Score 88.6;  DB 10;  Length 1523;
  Best Local Similarity   44.9%;  Pred. No. 8.5e-07;
  Matches  396;  Conservative    0;  Mismatches  479;  Indels    6;  Gaps    2;

Qy          4 GGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGC 63
Db        647 GGGGGCGGGGCGGGCGGGGGGCGGGGGNGGGGCGGGGCGGGGCGGGGCGGCGGGCGGGGC 706

Qy         64 GGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGC 123
Db        707 GGGCCGGGGGGCGGGGGGGCGGGGGNGCGGGGGGCGGGCGGGGCGCGGGGCCCGGGGGGG 766

Qy        124 CCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGAT 183
Db        767 GGGGGGGGGCGGGGGGCGGCGCGGGGGCGCGGGGGGGGGGGGCCGGGNGGGGGGGGGGG 826

Qy        184 GCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCAC 243
Db        827 GGGCCCGCGCGGCGCGGGGCCGCGGCCGGCGGGGGGGCGCGCGGGGGGGGGGGGCGCGGG 886

Qy        244 GGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATC 303
Db        887 GGCGGGGCNGGCGGCGGGGGGGGCGGCGGCGGCGGGGGGGGGGGCNGCGGGGGGGCGGGG 946

Qy        304 GCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGG 363
Db        947 GGGGGGGGGGCG--GCGGGGGGCGGGGCGGCGGGGGGGCGGCGGGCGGGGGCGGGGG 1004

Qy        364 CTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGC 423
Db       1005 CGGGCGCGGGCGGGGGGGGGGCGGGGGCGCGGGGGGGC----GGGGGCGGGGGGGGCGG 1060

Qy        424 GCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCGGCA 483
Db       1061 CGGCGGGGCGGGGGGCGGGCGGCGCGGGGCGGGGGGGGGCCGGGGGGGGGCGGGGCGGCG 1120

Qy        484 GAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCGCAG 543
Db       1121 GGCGGCGGCGGCGGGCGGGGGCGGGGGGGGCGGCGGGGGGCGGGGGGCGGCGCGGGGGCG 1180

Qy        544 CGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCACGAG 603
Db       1181 CGGGGGGGGGGGCGCGGCGCGCGGGGGCGGGGGGGCGGGCGGGGGGCGGGGGGCGGCGGG 1240

Qy        604 CTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAGCTT 663
Db       1241 GNGGCGGGGGGGGGGGCCGGGCGGGCGGCCGGGGGNGCCGGGCGGCGGCGGGGCCCGGGG 1300

Qy        664 GTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGCGGT 723
Db       1301 CGGGGGGGGGGGGGCGGCGGGGCGGGCGGCGCGGCGCGGCGGCCGCGGGGCGGGGGCGCG 1360

Qy        724 CACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCCGTG 783
Db       1361 CGCGGGGGGCGCGGGCGGCGGGCGGCGCGGCGGCGGGGGGGCGCGGGGGGGGGNGCCGGC 1420

Qy        784 ATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTCGCG 843
Db       1421 GCGGCGGGGGCGGGGCGGGGCGGCGGGGGCNGCGGCCGGCGGGGCGGCGGGCGCGGGGCG 1480

Qy        844 GGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGG 884
Db       1481 GCGCCGGGNCGCGCGCGCCGGCGGGGGCGGCGGGCGCGCGG 1521


RESULT 14
CNS0072Q
LOCUS       CNS0072Q                932 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066742
VERSION     AL066742.1  GI:4945205
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
```

```
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14B09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             9.8%;  Score 88.4;  DB 10;  Length 932;
  Best Local Similarity   34.2%;  Pred. No. 9.7e-07;
  Matches  158;  Conservative  114;  Mismatches  188;  Indels    2;  Gaps    2;

Qy          8 GCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGGCGCGGGC 67
              ||: |:::|: :  |: :: :  |: |  :  |  :  |::| :|| |   || :||:|
Db        453 GCSCTSSBGYCKCKGKCKKCBCGCKGGTKBCKGGGKGGCBKGCKGGCGGGGGGGVCGRGG 512

Qy         68 CGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAGCGCCCGG 127
               :  : |:| ||| |:|   |:|  ::  || | |:  | | | :::|  | ||:: |||
Db        513 GSGGSCGSTGGCGGGKGGCGCKGCTSK-GGCTGBGSTGCGGGGGSKSGSGGSGCSBGCGG 571

Qy        128 CAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAAGATGCCG 187
              : :||:|    :: : ::  :|: | ||::|:    |  |: |         :    ||
Db        572 GGSTKGGSGGGGGSBSBTKYTBKGKGGCGGSSGSCCSGGSCRGGGGGCGSCGGSGGCGCG 631

Qy        188 TGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAGCACGGCC 247
              |:::|||:     |:|| |||| ||:  :  |::  :  :|   ||| :: || :| ||
Db        632 GGSSSCGCSCSCGCGSGGCGCGCCGCSSGSGGCSSGGSGGSGSGCCGGCSSCGCCSGCCC 691

Qy        248 TCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCGATCGCGG 307
               :: |: |  || || :|| :  | || :|:   ::||  ||  |:|| |:|   ::||
Db        692 CSSGCSCCCCCGCGCGSCGGSCCG-CGCSCSGGCSSCGCSCCCSCSCCCCSCCCCSSCGC 750

Qy        308 ATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGGCGGCTCG 367
               :    |: ||::|| :| | : | :|| :|| ||  |  :  ||:|:: :|||:: :|
Db        751 SSSSGCGGSGGCSSGCGSCGCGSCGGSGCGSCGGCGGSCGSSSGCGSGSSGSGCGSSSSG 810

Qy        368 CCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCGCGCGCCC 427
              ::   | | ::|| :  ::   :: :|  | |:  |   |  :|: ||||: || :|:||
Db        811 SSCGGGSGCSSGCSSGGSSGCCSSGSGGCCCCSSGCGGCGCGSCSSCGCGSGGCCSGSCC 870

Qy        428 GCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCG 469
              || :| | ||||  |: ::  ||| :  || :|  :| | ::
Db        871 GCCSCGGCCGCGGCGSGSSCGGCGGSCSGCSSGCGSCCCGSS 912


RESULT 15
AG448243/c
LOCUS       AG448243               1780 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-332C01.TJ, genomic survey
            sequence.
ACCESSION   AG448243
VERSION     AG448243.1  GI:48091306
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1780)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1780
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-332C01.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             9.8%;  Score 88.2;  DB 10;  Length 1780;
  Best Local Similarity   45.2%;  Pred. No. 1e-06;
  Matches  405;  Conservative    0;  Mismatches  486;  Indels    6;  Gaps    3;

Qy          1 GGCGGGGGCCTTCCGGCTGGAGGGGTTCGACGAGCCCGAGCCGATCTGGGTGGAGATCGG 60
              |||| || || |  ||  |  | ||  || ||     | | |     | |  | |  |||
Db       1661 GGCGCGGCCCGTGGGGGGGCGGCGGCGCGCCGGCGGGGCGGCCGGGCGCGGCGCGNCCGG 1602

Qy         61 CGCGGGCCGCTGAGGTCGCGCGGGCTACGGGGCGACGCGGAGCGTCCGCGAGGCGACGAG 120
               |||| ||  | ||  | | ||||  |  || |   ||  || | |    || || |
Db       1601 GCCGGGNCGGCGCGGCGGGGGGGGCCGCCCGGGGGGCCGCCCCGGCGCCCGCCCGGCGGG 1542

Qy        121 CGCCCGGCAGAGGGCGATCCGGTCGTCGAGGTCGAGGCCGGGGAGCTCGCGCACGTAGAA 180
              |||||   |  ||||  || | ||  | |  ||        || |   |||| |
Db       1541 GGCCCGCGCGGCGGCGGGCCCGCCGGGGCGCGCGGCCGGCCGGGGGGGGCGCGCCGCCGC 1482

Qy        181 GATGCCGTGCCGCGCGATGAAGCGGAGCGCGGCCTCCCCCCGCAGATCGACGCGGACGAG 240
              |  |||| | || |||  |  || | |||| |||| |   ||| |       | | || |
Db       1481 G-GGCCGCGGCGGGCGGCGGCGCCGCGCGCCGCCTGCGGGCGCGGCGGCGGCCCGCCGGG 1423

Qy        241 CACGGCCTCGCCGTCGACGAGCTCCGCCTTGCCGTCCCTCAGCCGGACCGACGCCTCGCG 300
              |   |  ||| | ||    ||  | | | | |||  | |   |   |   |    |
Db       1422 CGGCGGGCCGCGGCCGGGCGGCCGGGGCCGCGGGCCGCGCGCCCGCGCGGGGCCGCGGGC 1363

Qy        301 ATCGCGGATCACGCGCCGCGGGCCGCACACGGACGCCGCGTCGCTCCACACCGCGGGCGG 360
              ||||  | |||| ||  ||| ||      |  |   ||  ||  | |||| ||
Db       1362 CGCGCGCGGCGCGCGGCGGCGGCGGCCGCGCCCGGGGGGCGCGGCCCGNCCGGCGGCGGG 1303
```

```
Qy        361 CGGCTCGCCGTAGAGGGCGCTGTACGCGGCCACGAGCTCGTCCCATGTCGCCTCGCGGCG 420
              |  |  ||  |  | |   |  |||  | |  || ||  ||  | |||  |   |||
Db       1302 GNGGGCCGCGGGGGCGCCCGGGGCGGCGCGCGCCCGCGCGCGCCCGGGCGCGGCNGCGCG 1243

Qy        421 CGCGCCCGCGGCCGGCGCGTTGCTCGGCGCGTGGTGCAGGAAGCGCCCGAAGAAGCGCCG 480
              ||||  |  || ||||||      | |||    | |  ||  | || ||  |  ||||
Db       1242 GGCGCGGGGCGCGGGCGCG----CCCGCGGCGCGGGGGGGGCGGGCGCGGGGGGCGCGN 1187

Qy        481 GCAGAACTCGGCGTATTCGAGCGTGAAGAGGGCGAACTGGTGCCAGGCCTCGTCGACGCG 540
              |       | |||     |   | |  | ||||     || | |  |   ||    ||||
Db       1186 GGCCNGGCCCGCGGCGGGGCCGGCGCGGGGGGCCGGGGGGCGGCGCGGGGCGCGCGCGCG 1127

Qy        541 CAGCGAGAACATCGGATAGGCGCGGGAGCGGTCGATCTCGACGAGCCAGAGATAGCGCAC 600
              |   | |  |   |     |  ||| | | || ||  |    || ||  | |   || | |
Db       1126 CGCNGCGGGCGGGGCCGCGCGGCGCGCGGGGCCGCGCCGNGCGCGCGGGGGGCCGCCCGC 1067

Qy        601 GAGCTCCCGGAACAGCGCCTCTGCCTCCTCCCGGGTGGCCACGGTCTTGTTCATGAGCAG 660
              | |  | | |    |  | | |  | |  |   |||  | | |  |  |  |
Db       1066 GCGGGCGCCGCGGGGGCGCCCCGGC-GCGCGGGCCGGGCGGCCGGCGGGGGCCGGGCGGC 1008

Qy        661 CTTGTCGATCACGAAGGGCGCCCGGTAAGCGAAGAGATCAGGCGTCCTGCGCTGGGTCGC 720
              ||    ||  || ||| |||   |||    |  | ||||  |   |  ||   ||
Db       1007 GGGCGCGCCGCCGNCGGCCGCGCGGCCCGCGCGCGGGGCGGGCGGGCGNGGGCGGCCGGC 948

Qy        721 GGTCACGATGTCCGTTTGCATGGGTCAGTTCTCCTGGGCTTCGAGCGGCTGAAAGGTGCC 780
              |       |   |   ||  |||   |  | || | |      ||| |     |  ||
Db        947 CGCGGGCGCGCGGGGCGGCGCGGGCGCGGGCGCCGGCGGCNGNGGCGCCGCGCCGCGGCG 888

Qy        781 GTGATCGACGAGCGCGCGGGCGAGCGCGAGCTGCTCGGCCTCGGCGAGGCCGGGGATGTC 840
              |     | || || ||  |  | |||   |     ||  | || |  |   ||
Db        887 GCCGCGGGCGCGCCCGGCGCGGGGCGNCCGGGCGCCGCGCCCGCCCGGCGGGGCNCCCGN 828

Qy        841 GCGGGGGCGGAGCTCGCGGGCGGCGGCGAGCGCGCGGAGCGCGGGCGCGGCCCACGC 897
              |  |||  || |  |||||  ||||    | | | |  |  | |||||  ||||  | |
Db        827 GGCGGGCGGGCGNGCGCGGCGGGCGCGCGGGGGGGGCGGNCCCGGCGCCGCCCGCCC 771
```

Search completed: April 28, 2006, 12:17:52
Job time : 4466.48 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 14:40:19 ; Search time 1634.62 Seconds
(without alignments)
10258.556 Million cell updates/sec

Title: US-10-729-802-1_COPY_77000_77294
Perfect score: 295
Sequence: 1 gcgagggcggtctgcgcgtc.....ggggccgagcaggtgcagcg 295

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	295	100.0	84868	1	DQ013294	DQ013294 Polyangiu
2	72	24.4	110000	1	BX571965_03	Continuation (4 of
3	72	24.4	110000	1	CP000010_34	Continuation (35 o
c 4	70.4	23.9	73882	6	CQ870950	CQ870950 Sequence
c 5	70.4	23.9	89713	1	AJ605139	AJ605139 Actinopla
c 6	70.4	23.9	89976	1	AJ632270	AJ632270 Actinopla
7	70	23.7	167624	14	AC143286	AC143286 Macaca mu
8	69	23.4	169214	8	AC092720	AC092720 Homo sapi
9	67.8	23.0	1279	10	PM2H12G	AL684840 Penicilli
c 10	67.8	23.0	65351	14	AC139773	AC139773 Homo sapi
c 11	67.8	23.0	156842	14	AC026284	AC026284 Homo sapi
12	67.2	22.8	296500	1	SCO939128	AL939128 Streptomy
c 13	67	22.7	143360	14	AC166628	AC166628 Sorex ara
c 14	66.8	22.6	1094	10	PM7G11B	AL685196 Penicilli
c 15	66.4	22.5	219952	14	AC084804	AC084804 Mus muscu
c 16	66.2	22.4	176386	14	CR855271	CR855271 Danio rer
17	66	22.4	885	10	PM7F8G	AL685189 Penicilli
c 18	66	22.4	67246	14	AC105035	AC105035 Homo sapi
c 19	66	22.4	72645	14	AC112672	AC112672 Mus muscu
20	66	22.4	110000	14	AC164423_7	Continuation (8 of
21	66	22.4	138203	1	AY310323	AY310323 Streptomy
c 22	66	22.4	166634	14	AC141736	AC141736 Apis mell
23	65.6	22.2	150715	13	AY714813	AY714813 Cercopith
c 24	65.6	22.2	150715	13	AY714813	AY714813 Cercopith
25	65.6	22.2	301399	1	AE017233	AE017233 Mycobacte
c 26	65.4	22.2	207592	14	CR392332	CR392332 Danio rer
27	65	22.0	136753	14	AC149968	AC149968 Strongylo
c 28	64.6	21.9	8407	1	SEU82823	U82823 Saccharopol
c 29	64.6	21.9	48000	1	POL505006	AJ505006 Sorangium
c 30	64.4	21.8	1338	6	AR622202	AR622202 Sequence
c 31	64.4	21.8	20113	6	AR619778	AR619778 Sequence
c 32	64.4	21.8	167624	14	AC143286	AC143286 Macaca mu
c 33	64.2	21.8	3692	1	AY701054	AY701054 Streptomy
c 34	64.2	21.8	110000	14	AC164423_7	Continuation (8 of
c 35	64	21.7	1279	10	PM2H12G	AL684840 Penicilli
36	64	21.7	48144	14	AC069166	AC069166 Homo sapi
37	64	21.7	180801	9	AC125125	AC125125 Mus muscu
c 38	64	21.7	216389	9	AC122892	AC122892 Mus muscu
39	63.8	21.6	82746	1	AF453501	AF453501 Actinosyn
40	63.8	21.6	207592	14	CR392332	CR392332 Danio rer
41	63.6	21.6	3730	9	AK122388	AK122388 Mus muscu
42	63.6	21.6	3959	6	AX305257	AX305257 Sequence
43	63.6	21.6	3959	9	MUSNFH	M35131 Mouse neuro
44	63.6	21.6	3960	9	MMNFHGE	Z31012 M.musculus
c 45	63.6	21.6	11305	15	OSA535061	AJ535061 Oryza sat

ALIGNMENTS

RESULT 1
DQ013294
LOCUS DQ013294 84868 bp DNA linear BCT 14-MAY-2005
DEFINITION Polyangium cellulosum strain So ce12 disorazole biosynthesis gene cluster, complete sequence.
ACCESSION DQ013294
VERSION DQ013294.1 GI:63099932
KEYWORDS .
SOURCE Polyangium cellulosum
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 84868)
AUTHORS Julien,B., Reid,R., Fehd,R. and Gramajo,H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .84868
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce12"
/db_xref="taxon:56"
misc_feature 1. .84868
/note="disorazole biosynthesis gene cluster"
CDS complement(188. .1666)
/note="ORFu1; ORFu1p; MHS family; similar to Pseudomonas putida KT2440 PP2569"
/codon_start=1
/transl_table=11
/product="metabolite MFS transporter"
/protein_id="AAY32956.1"
/db_xref="GI:63099933"
/translation="MRSAPGPPPRGRRRAEARPQPSRRVLGIFHRGTAPMAQLSPETL
PVTGTPRAREAEREVVSSSTLWKVIFASAGGTMIEWYDFYIFGSLAAIISTHFFPKDN
PTAGFLLTLGTFATGFVVRPFGALVFGRIGDVVGRKYAFLVTLLLMGGATSIIGILPT
YDKIGLAAPLALVLIRLVQGLALGGEYGGAAVYVAEHAPDEKRGFYTSFIQITATLGL
FVSLAVILGTRILLGEEAFKEWGWRVPFLLSILLVAVSFYVRARLKESPIFQKLKASG
KTSSAPIKDSTRNWRVILIALFGATAGQGVVWYTGQFYALYFLQNTLKVPFITATIVV
AVALALGMPFFVLFGALSDRIGRKKIMLAGNLLGAIGFYPIFWAMKQNADSPAVLTAL
VFVLVLFVTMVYGPIAAYLVEAFPAKVRYTSLSIPYHVGNGVFGGMLPLIASSVTVAT

```
                GNIYAGLAYPIVVALITVVVGGIYLKERSDVRLWDEVDREAA"
CDS             complement(1992. .3440)
                /note="ORFu2; similar to Pseudomonas putida KT2440 PP4696;
                putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32957.1"
                /db_xref="GI:63099934"
                /translation="MSRILLIEDEIVIRAELRRLLSRHGHAVVEAGSIAEAEAGHDLS
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVEHHLASARSARGALPEVAPHVPSSAAPPQSVPPPQSVPPPQSVPPPQSVPP
                APDDDELEGMVGGSAAMREVFAQVRKVAPSEATLLVVGEGGTGKETLSRAVHRLSPRR
                QQPFVLVQCAAIPDAQIEGELFGYEAGAAAGALTARAGLCEAAEGGTLLLDEIGELPM
                KVQARLLRFLQTGEVRRVGSTRARRADVRLIAATHRDLGAMVREGSFRSDLYFRLRVL
                EIRLPPLRERPGDAVRIAESLLKRRTRMGQAPLALSAEARAAIAAAAWPGNVRELENA
                VHRAVVLAGGSAIERAHLGLDDPPRPSPSAPPAPQREATPPSDSLEGYFRRFVLEHQD
                ELTETEIARRLGISRKALWERRQRLGIPRPRS"
CDS             3587. .5548
                /note="ORFu3; ORFu3p; similar to Leptospira interrogans
                AAN50474"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32958.1"
                /db_xref="GI:63099935"
                /translation="MHKTTLLRAATAEHAPDRAATRHRGMDSVPALPFVARPRRMLPI
                VARRHRLAQRARSPPAVPRPARALTRHGSAGAPRLPPPRPHVPRRSCSRISTRILPSM
                LEHPCLRQPQRFDIRFVFDPDFYRHIDAQYTDIPGGCLRSTFGPSLLTPPAPAAARDR
                YATPLCPSSAPTQRGQVAWTSRGRSRSLVRVTATLRTTSIVAFLRRSSTDAEEDDEKR
                SKTAAVRAGIFLCLTLTAAAPACTPFRSKECAGIALQRPVEALWGRRFRGNGSVRGVA
                IDATGTIAMTGWYTGTLDLDDGPRAVIDVDGSNDMFVVKLDAQGKMLWTRHFGGARMQ
                VGNDVASDREGNVIAAGKVRGSIDFGTGTLHADDLKSSAVVVKLDPLGHTRWAKLFQG
                AERAESVAVDASGDVYVATAVRGGLSVVKLDPGGQSVFDVHLGHLEGFRGNRARMVAD
                GHGHLVVVWNHIAEGYVLLGDELSRHDYRERRTSVAKLDASGHVLWSHVLGEDSPRDV
                AIGARGDVVIVGDAGDDETAAGFATKLAADGGHLWTKGLGLRMWPRGVAVDARGDLVI
                SGITSGEVDVCGRDRTGPEQITWLAKLDPNGRFVWARRVELGAWGGGEGVAIHPAGGV
                VVAGFGATGPLRVMETLPEQYLFLTRVRP"
CDS             complement(5702. .7231)
                /note="ORFu4; ORFu4p; similar to Burkholderia fungorum
                hypothetical protein in ZP_00033665"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAY32959.1"
                /db_xref="GI:63099936"
                /translation="MRMTLMKSPTGTVLLALALALAGCASSGDDGPSQGGGAPGASSS
                AAAGSTSGTGSTSGTGSAAGSTSAAPGGTGGEGGAGAGGAGGEGGAGVGGAGGSAGGE
                PRIPWDWVGIVGTGQSLSVGATPITTTTQPFKNLKLDLGSASVPPWGPDHSALSMVPL
                VEPIRSVDIMTGRPAYPANIYGETPHSAMANQISFLVQEAAGRDYVSVHTVVGESGQA
                MSVIEKGATDSGTTGRAYAATLFEAEAIARLARAAGKTYGIGAIFMTHGETDWASATY
                QDDLVQLWNDYNADLAALTGQTGGKIPMFVSQQHGFPFSANQRPASTLAQWRVGLEHP
                GDIVCTGPKYHLPAASDGLHLTAAGYKQLGEKYGQIYYEKVVMGRDWQPLQPTGAERS
                GRVITVHFHVPVPPLVWDDGMPAPRLWPDGRGFEVRQANTKVDITSVEISGSSVRITC
                ASDLPASDVTVGYALTTDSTPMPGGFRAWGHLRDSDPFVGWATHVAQPNYAVAFEMPV
                P"
CDS             complement(7561. .8931)
                /note="ORFu5; ORFu5p; similar to Pseudomonas putida KT2440
                PP4696; putative nitrogen regulation protein NR(I)"
                /codon_start=1
                /transl_table=11
                /product="putative regulatory protein"
                /protein_id="AAY32960.1"
                /db_xref="GI:63099937"
                /translation="MSRILLIEDEIVIRAELRRLLARHGHAVVEAGSIAEAEAEHELC
                AFDLVLADVRLPGAPGTELVGRCGSAPVVLMTSYATVKAAVEAVKRGAADYLSKPFDH
                DELVALVERHLASVRPEAAPPAPLAAPPAPPAPDGDELEGMVGGSAAMREVFAQVRKV
                APSEATLLVVGEGGTGKETLARAVHRLSPRRQQPFVLVQCAAIPDAQIEGELFGYEAG
                AAAGALAARAGLCEAAEGGTLLLDEVGELPMKVQGRLLRFLQTGEVRRVGSTRARRAD
                VRLIAATHRDLGAMVREGSFRSDLYFRLRVLEIRLPPLRERPGDAVRIAESLLKRRTR
                VMGQAPLALSAEARAAIAAAPWPGNVRELENAVHRAVVLAGGSAIERAHLGLDDPPRP
                PPPVSSASLRDGAPPSDSLEGYFRRFVLEHQDELTETEIARRLGISRKALWERRQRLG
                IPRPRS"
CDS             complement(8928. .11939)
                /note="ORF1; ORF1p; similar to Pseudomonas putida KT2440
                HisK (PP4695) in GenBank Accession Number AAN70268"
                /codon_start=1
                /transl_table=11
                /product="putative sensory box histidine kinase"
                /protein_id="AAY32961.1"
                /db_xref="GI:63099938"
                /translation="MGVLLFATVAWLGVLAAIALAAQRARRPRALARSPLVLALSLGV
                YASSWTYYGGVGFAQRHGLVFLAVYLGPTLACMATPLLWARVLRITREQRLSSLTDLL
                AFRFQSPAAGALVAAFALVASLPYVAQQFRATVQSVRVLSGPGHEEAVGAVFCALLAL
                FTAPFGARPSAEQGREDGLAAAVAFESAAKLAVLAAVGLAAVLGAFGGLGGLAGWLGE
                HAAEVEALGRHVREDSGWTSMLVISFAAAFLLPRQWHMAFASGAGDRALATAAWAFPL
                FLLALNLPVLPVLWAGTRLAPAEPADSYVLLVPQLLGQRWLALAAFIGGISAASAMAM
                VTAVALANDSATHLLLPLARRTLRRNVYARVLWLRRVLLAAILVGGWLFHRAQPHHGL
                LVESGIVSFVAFAQFLPGLLAALFWERATRAGFMAGLLAGIATWTATCLAPLLPGVPG
                IDPAWLALLRGGAGAPAGAAPIDPWSLPTFVSLAVNGATLGIVSLLRPPAREEREAAR
                LCTRREIGPSVERATVASSVAALEDRLARLLGREAAAIEVTRARIELDISRDETRPTE
                LHALQERLHRNLSPLVGPVLARTAAEAEASAPVASGGASGALTEELRWLERRLARGET
                RLVGAARELDLLRRWLRSVLAQLPLGVCAVGSDGAISLVNRRLEALVGAEERDLVGAP
                LDRLPAPFAELFARAARAASAGEAGAGAEEIEVLVAGRARLLRVHRAALDPLDARAAS
                GGVVLLVEDRTEQRALEAQLAHRDRLAQIGRLAAGVAHEVGNPLTGIACLAQNLRAEA
                DSDEIRARAGLILGETARIEAILRSLLDHSRTGAVPSPGEIPRARVGRVSLAEVVAEA
                DRLIRLDRTAPHVRVEQHCPESLVVIGDRRELVQVFVNLLANARDASEPGATVAVRGW
                IEAGQIVVEVEDHGSGIPEELLSAVLEPFVTTKRDPSGTGLGLPLSRSIIEDHGGALA
                LRSAVGQGTTVIVRLPLAGEAGARRGREVEA"
CDS             complement(12405. .13379)
                /note="ORF2; ORF2p; similar to family of known or putative
                phosphotransferases, including macrolide
                2-phosphotransferases; similar to Bacillus subtilis YcbJ,
                Bacillus halodurans MphB, Escherichia coli plasmid pTZ3273
                MphB and Staphylococcus aureus plasmid pSR1 MphBM"
                /codon_start=1
                /transl_table=11
                /product="putative macrolide kinase"
                /protein_id="AAY32962.1"
                /db_xref="GI:63099939"
                /translation="MTEHHELSSTASSTASSDFAPLDSPGALLDAARREGLRIVAPRA
                DFDQSGLDFLVVHARDEEGVPWVVRTPRRLAVVKAARVEARVLALVRKHLPVAVPDWR
                VHTDQVIAYPRLGDVPAVSFDPEAGVRWNHVDPAAPPAAFIDSMADALAALQRIDAAA
                IEAAGVPVKSIDDERATLARAMDETRSTLSPPEAVWARWQRWLADDARWPRHVALVHG
                DLHPGHMLVREDGRLTGILDWTEAHVGDPSSDFALFAGAFGKAVLDAVIARFEAAGGP
                AWPGLAAHAAERWAAWPALAAKWGIDMGNEAVIEHARSHLEKIAAESG"
CDS             13368. .14663
                /note="ORF3; ORF3p"
                /codon_start=1
                /transl_table=11
                /product="putative Ser/Thr protein kinase"
                /protein_id="AAY32963.1"
                /db_xref="GI:63099940"
                /translation="MFGHDRGPLRAPPAGRRAWKGEDSRRELTTDQACMACILQAAGR
                ESTRPASSTCRRKTGHGQAPARSPRRPGGAGEAGRTRMPGSAQPEKCHGPRFSRGVPR
                RNPLRARGAAARGDSIARMGIDEELAEQRIGTRIGPWSVERVLGVGGMASVYYCRRDD
                GCVAAVKLLHPELASIEEVRKRFLREGPIGSALAAVAPLCEGLPQVIEAGEADGAAYM
                AMEMLEGETVFDRMVRHGTLPVGQVIALAERVLDVLDVAHAHGIVHRDLKPENLHIGN
                DGRVRVLDFGLARVLDPLLEDVAGVPEMTKTSTGVSIGTDDYMAPEQALGLIREIDGR
                TDLFGLGATMFRLLAGRTIHGNLEDAHLLIAAATEKAPPLAQHAPAAPPGLCAVVDRA
                LAFLKQERYPDARTMRADLAAVRAGREPPYATAAARGRA"
gene            15731. .33766
                /gene="dszA"
CDS             15731. .33766
                /gene="dszA"
                /note="PKS (polyketide synthase) subunit modules 1-4a"
                /codon_start=1
                /transl_table=11
                /product="DszA"
                /protein_id="AAY32964.1"
                /db_xref="GI:63099941"
                /translation="MEQDAIAIIGVACRFPGSPDYGRYWQLLERGEHAILEIPPGRWD
                PRAHYSPDFNKPGKSISKWCGLIDDIASFDHRFFNVSEREAKSMDPQQRLLLEEAWRC
                IEDSGVPLEQLRARKTSIYVGFMATDYHQESAAPGRPVDSYAALGSYGSILANRVSYT
```

```
                     LGLRGASIAIDAACASSLVALHEARRALQRGESEFALAAGVSLNFHPWKYVSFSKSRM
                     LSPDGLCKTFDADANGYVPGDGVGVLLLHPLAKAIAAGCHVYGVVAGSAVNHTGTARS
                     ITAPRVAAQRDVILEAYEDAGWNPETVTYVEAHGTGTSLGDPIEVEALTQAFRRYTTA
                     RQRCAIGSVKSNIGHLEAAAGVAGVIKVLMMLKHRVIPRTLHVQTLNPLIRFEETPFV
                     VATRAMEWRAEGGEPLRAGVSSFGFGGANAHVLISEHGGARREPRPRGELRGPRGAAP
                     RGETAGAPAEDGPLARAEELPSQQEDAAADEREGTVFLLSARSASSLSRAVRRWEAFV
                     DDPLVKAGLATSLRDICATLAAGRQSFEHRHGFYIDDERDLRRLLKEPPARLEKTRPP
                     RWVTRFGALALGQGRPAVRLLGARRLLDPHLDRIRRCLEELGIEHQDLRTYRQDGDPG
                     RQELPYAFLFAHAYVSALADLGFTPYATSGEGHGIWLALAQSGVLPLNEIVSVLSGAG
                     ELQRLSPRRPRLPLFDPIHSTYLMPYLLDAGYVRALVEGLAVPAATLRDLLARARLLL
                     RAQFTFKKFLSEWSPALQALGTTPERLLEEELPASDARASLIALIAQSCVRKLNRRWQ
                     LTDAPSSGDPRFDELVDLVVDGLLPREALVQLALGDRADLHEIAGTLHRRQDLLDLSQ
                     PYGILRRRSERLDPSEIDDFSGWIRQIAGLEAPGLPPEEGVAFLELGRVARRAQRAPG
                     PDLSVPALDSPLQLVALRLWLEGTDIRWGELFPEGSFAKIPLPGYAFDQAQFWLPAAR
                     EGTSPPEDARDDADARHAAVAPHGAADRAERPSIPVDRLIADHVIQGRAIVPGALMVE
```

Query Match 100.0%; Score 295; DB 1; Length 84868;
Best Local Similarity 100.0%; Pred. No. 3.6e-18;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84574 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 84633

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84634 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 84693

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84694 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 84753

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84754 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 84813

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84814 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 84868
```

RESULT 2
BX571965_03
WPCOMMENT
Sequence split into 41 fragments LOCUS BX571965 Accession BX571965

Fragment Name	Begin	End
BX571965_00	1	110000
BX571965_01	100001	210000
BX571965_02	200001	310000
BX571965_03	300001	410000
BX571965_04	400001	510000
BX571965_05	500001	610000
BX571965_06	600001	710000
BX571965_07	700001	810000
BX571965_08	800001	910000
BX571965_09	900001	1010000
BX571965_10	1000001	1110000
BX571965_11	1100001	1210000
BX571965_12	1200001	1310000
BX571965_13	1300001	1410000
BX571965_14	1400001	1510000
BX571965_15	1500001	1610000
BX571965_16	1600001	1710000
BX571965_17	1700001	1810000
BX571965_18	1800001	1910000
BX571965_19	1900001	2010000
BX571965_20	2000001	2110000
BX571965_21	2100001	2210000
BX571965_22	2200001	2310000
BX571965_23	2300001	2410000
BX571965_24	2400001	2510000
BX571965_25	2500001	2610000
BX571965_26	2600001	2710000
BX571965_27	2700001	2810000
BX571965_28	2800001	2910000
BX571965_29	2900001	3010000
BX571965_30	3000001	3110000
BX571965_31	3100001	3210000
BX571965_32	3200001	3310000
BX571965_33	3300001	3410000
BX571965_34	3400001	3510000
BX571965_35	3500001	3610000
BX571965_36	3600001	3710000
BX571965_37	3700001	3810000
BX571965_38	3800001	3910000
BX571965_39	3900001	4010000
BX571965_40	4000001	4074542

Continuation (4 of 41) of BX571965 from base 300001 (BX571965 Burkholderia pseudomallei

Query Match 24.4%; Score 72; DB 1; Length 110000;
Best Local Similarity 53.6%; Pred. No. 75;
Matches 150; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

```
Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||||| ||||    ||||  |   || ||   |    ||  |||| | ||| ||  ||
Db      37512 GCGAGCGCGGCGGCCGCGCGGGCTTCGTCGCCCGCCGCGGCCGCGAGCGCGACGTTCCAG 37571

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                 ||  |||  |||| |||| ||  |||  ||| |  |||||||  |  |||     |
Db      37572 CCGCGCCGCGCGAAGCCGACGGCGAGCGCGCGGCCGACGGCCGCGGCGTCAGGCGCGGGC 37631

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | |    | | |   || || ||||  |   |     |||||| |||| ||| | |  |
Db      37632 CGCGCGGCCGGCGGCGCGCTCGGCGCTTGCGCCTGCTGCTCGCCGGCGCGCGGGGGATGT 37691

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              |||   |     ||   | | |  |||    ||| ||||  |||| | | |||||  ||  |
Db      37692 GCCGGCGCACCGGCTGGCGCCGATGCGCCGGGCTGCGCATCGCCCGGCCGGCCGGCGCCG 37751

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGC 280
              | |   | ||||  ||      ||    | |||  || ||
Db      37752 ATGGCTTGCCGCGCGCCCGCCCGCGCCGCCTCGGCGGCGC 37791
```

RESULT 3
CP000010_34
WPCOMMENT
Sequence split into 36 fragments LOCUS CP000010 Accession CP000010

Fragment Name	Begin	End
CP000010_00	1	110000
CP000010_01	100001	210000
CP000010_02	200001	310000
CP000010_03	300001	410000
CP000010_04	400001	510000
CP000010_05	500001	610000
CP000010_06	600001	710000
CP000010_07	700001	810000
CP000010_08	800001	910000
CP000010_09	900001	1010000
CP000010_10	1000001	1110000
CP000010_11	1100001	1210000
CP000010_12	1200001	1310000
CP000010_13	1300001	1410000
CP000010_14	1400001	1510000
CP000010_15	1500001	1610000
CP000010_16	1600001	1710000
CP000010_17	1700001	1810000
CP000010_18	1800001	1910000
CP000010_19	1900001	2010000
CP000010_20	2000001	2110000
CP000010_21	2100001	2210000
CP000010_22	2200001	2310000
CP000010_23	2300001	2410000
CP000010_24	2400001	2510000
CP000010_25	2500001	2610000
CP000010_26	2600001	2710000

```
     CP000010_27       2700001   2810000
     CP000010_28       2800001   2910000
     CP000010_29       2900001   3010000
     CP000010_30       3000001   3110000
     CP000010_31       3100001   3210000
     CP000010_32       3200001   3310000
     CP000010_33       3300001   3410000
     CP000010_34       3400001   3510000
     CP000010_35       3500001   3510148
Continuation (35 of 36) of CP000010 from base 3400001 (CP000010 Burkholderia mallei ATCC

  Query Match            24.4%;  Score 72;  DB 1;  Length 110000;
  Best Local Similarity  53.6%;  Pred. No. 75;
  Matches  150;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||||| ||||    ||||  |   || ||   |    ||  |||| | ||| ||  ||
Db      92599 GCGAGCGCGGCGGCCGCGCGGGCTTCGTCGCCCGCCGCGGCCGCGAGCGCGACGTTCCAG 92658

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                 ||  |||  ||||  |||| ||  |||  ||| |  |||||||  |  |||    |
Db      92659 CCGCGCCGCGCGAAGCCGACGGCGAGCGCGCGGCCGACGGCCGCGGCGTCAGGCGCGGGC 92718

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | |    |  | |   || || ||||  |    |    ||||||  |||| ||| | |  |
Db      92719 CGCGCGGCCGGCGGCGCGCTCGGCGCTTGCGCCTGCTGCTCGCCGGCGCGCGGGGGATGT 92778

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              |||   |    ||  | | |  |||   ||| ||||  |||| | | |||||  ||  |
Db      92779 GCCGGCGCACCGGCTGGCGCCGATGCGCCGGGCTGCGCATCGCCCGGCCGGCCGGCGCCG 92838

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGC 280
                | |   | ||||  ||     ||   | |||  || ||
Db      92839 ATGGCTTGCCGCGCGCCCGCCCGCGCCGCCTCGGCGGCGC 92878


RESULT 4
CQ870950/c
LOCUS       CQ870950               73882 bp    DNA     linear   PAT 27-SEP-2004
DEFINITION  Sequence 1 from Patent EP1460085.
ACCESSION   CQ870950
VERSION     CQ870950.1  GI:52745148
KEYWORDS    .
SOURCE      Actinoplanes teichomyceticus
  ORGANISM  Actinoplanes teichomyceticus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Micromonosporineae; Micromonosporaceae; Actinoplanes.
REFERENCE   1
  AUTHORS   Donadio,S., Sosio,M. and Bianchi,A.
  TITLE     Genes and proteins for the biosynthesis of the glycopeptide
            antibiotic teicoplanin
  JOURNAL   Patent: EP 1460085-A 1 22-SEP-2004;
            Vicuron Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .73882
                     /organism="Actinoplanes teichomyceticus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1867"
     misc_feature    652. .1995
                     /note="ORF1; positive strandedness"
     misc_feature    2017. .3234
                     /note="ORF2; positive strandedness"
     misc_feature    3227. .4264
                     /note="ORF3; positive strandedness"
     misc_feature    4264. .4872
                     /note="ORF4; positive strandedness"
     misc_feature    5634. .6173
                     /note="ORF5; negative strandedness"
     misc_feature    6286. .7377
                     /note="ORF6; negative strandedness"
     misc_feature    7385. .8065
                     /note="ORF7; negative strandedness"
     misc_feature    8924. .10105
                     /note="ORF8; positive strandedness"
     misc_feature    10597. .16827
                     /note="ORF9; positive strandedness"
     misc_feature    16824. .20009
                     /note="ORF10; positive strandedness"
     misc_feature    20053. .32256
                     /note="ORF11; positive strandedness"
     misc_feature    32276. .37873
                     /note="ORF12; positive strandedness"
     misc_feature    37886. .38095
                     /note="ORF13; positive strandedness"
     misc_feature    38401. .39222
                     /note="ORF14; positive strandedness"
     misc_feature    39268. .41049
                     /note="ORF15; positive strandedness"
     misc_feature    41343. .43289
                     /note="ORF16; positive strandedness"
     misc_feature    43373. .43585
                     /note="ORF17; positive strandedness"
     misc_feature    43696. .44871
                     /note="ORF18; positive strandedness"
     misc_feature    44894. .46048
                     /note="ORF19; positive strandedness"
     misc_feature    46038. .47234
                     /note="ORF20; positive strandedness"
     misc_feature    47484. .49004
                     /note="ORF21; positive strandedness"
     misc_feature    49166. .50344
                     /note="ORF22; positive strandedness"
     misc_feature    50461. .51687
                     /note="ORF23; positive strandedness"
     misc_feature    51833. .52804
                     /note="ORF24; positive strandedness"
     misc_feature    52948. .54540
                     /note="ORF25; positive strandedness"
     misc_feature    54757. .56553
                     /note="ORF26; positive strandedness"
     misc_feature    56924. .57979
                     /note="ORF27; positive strandedness"
     misc_feature    58746. .59735
                     /note="ORF28; positive strandedness"
     misc_feature    60477. .62855
                     /note="ORF29; positive strandedness"
     misc_feature    63036. .64157
                     /note="ORF30; positive strandedness"
     misc_feature    64154. .64819
                     /note="ORF31; positive strandedness"
     misc_feature    64816. .66117
                     /note="ORF32; positive strandedness"
     misc_feature    66146. .66952
                     /note="ORF33; positive strandedness"
     misc_feature    66991. .68352
                     /note="ORF34; positive strandedness"
     misc_feature    68349. .68945
                     /note="ORF35; positive strandedness"
     misc_feature    69083. .70180
                     /note="ORF36; positive strandedness"
     misc_feature    70266. .71321
                     /note="ORF37; positive strandedness"
     misc_feature    71318. .72412
                     /note="ORF38; positive strandedness"
     misc_feature    72548. .73300
                     /note="ORF39; positive strandedness"
ORIGIN

  Query Match            23.9%;  Score 70.4;  DB 6;  Length 73882;
  Best Local Similarity  54.7%;  Pred. No. 1.2e+02;
  Matches  162;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    1;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              | |  | |||  ||     | || |||| ||      ||||||  |    ||| ||||
Db      35827 GTGCACCGTCACCGACAGCACGTGCTCCTTCTCCGACAGCGCGAAGAGCTCGCACCGCCA 35768
```

```
Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCG---CGGCCC 119
              || |  | |   |  | || ||  |||  || ||    || ||||   |||    ||
Db      35767 CGGCATGTCCCGGGTGAGGTCGAACAGCTGCTCGCCCCGCTCGGCCAGCCGTGCGGGGAG 35708

Qy        120 CCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAG 179
                |  ||||  |||   || | ||| | | || ||  || ||  | ||||   |||  |
Db      35707 GTCCTCCTCGGTGGCCGGGACCGGCGTCAGCCGGACCGGCGCGGCGTCGTGGATGTGCTG 35648

Qy        180 CGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGAT 239
                 ||| | |||||||  | |||  ||  | | ||  |    ||  | || |  |||
Db      35647 GTGGACGGTCTGCGCGTCGCCCGGGAAGGTGGTCCGCAGGATCTCGTGCCGGGCCGCGAC 35588

Qy        240 GGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |  || || |  ||| |||||| | |   ||||  |  ||||| ||||| ||||||
Db      35587 GTCGCGCAGCGCCGCCTCCAGGGCGGCCACGTCCAGCCGGCCGCGCAGGCGCAGCG 35532
```

RESULT 5
AJ605139/c

```
LOCUS       AJ605139               89713 bp    DNA     linear   BCT 15-APR-2005
DEFINITION  Actinoplanes teichomyceticus gene cluster for biosynthesis of
            glycopeptide antibiotic teicoplanin.
ACCESSION   AJ605139
VERSION     AJ605139.1  GI:40713132
KEYWORDS    ABC transporter; acyl-CoA ligase; AfsR type regulator; AraC-family
            transcription regulator; DpgA protein; DpgB protein; DpgC protein;
            DpgD protein; glucanase; GtfA protein; GtfB protein; GTP
            cyclohydrolase; halogenase; HmaS protein; Hmo protein; HpgT
            aminotransferase; LuxR family regulator; mannose-6-phosphate
            isomerase; mannosyltransferase; MbtH-like short polypeptide;
            membrane ion antiporter; membrane zinc protease; MerR family
            regulator; murF-like protein; non-haem hydroxylase; non-ribosomal
            peptide synthetase; Oxy protein; OxyA protein; OxyB protein; OxyC
            protein; phospho-2-dehydro-3-deoxyheptonate aldolase; RCC1 repeats
            protein; response regulator; short chain dehydrogenase; StrR family
            regulator; tcp1 gene; tcp10 gene; tcp11 gene; tcp12 gene; tcp13
            gene; tcp14 gene; tcp15 gene; tcp16 gene; tcp17 gene; tcp18 gene;
            tcp19 gene; tcp2 gene; tcp20 gene; tcp21 gene; tcp22 gene; tcp23
            gene; tcp24 gene; tcp25 gene; tcp26 gene; tcp27 gene; tcp28 gene;
            tcp29 gene; tcp3 gene; tcp30 gene; tcp31 gene; tcp32 gene; tcp33
            gene; tcp34 gene; tcp35 gene; tcp36 gene; tcp37 gene; tcp38 gene;
            tcp39 gene; tcp4 gene; tcp5 gene; tcp6 gene; tcp7 gene; tcp8 gene;
            tcp9 gene; tei1 gene; tei10 gene; tei11 gene; tei12 gene; tei13
            gene; tei14 gene; tei2 gene; tei3 gene; tei4 gene; tei5 gene; tei6
            gene; tei7 gene; tei8 gene; tei9 gene; TetR-family transcitpional
            regulator; thioesterase; transcriptional regulator; transmamebrane
            sensor kinase; transposase; VanA protein; VanH protein; VanX
            protein.
SOURCE      Actinoplanes teichomyceticus
  ORGANISM  Actinoplanes teichomyceticus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Micromonosporineae; Micromonosporaceae; Actinoplanes.
REFERENCE   1
  AUTHORS   Sosio,M., Kloosterman,H., Bianchi,A., de Vreugd,P., Dijkhuizen,L.
            and Donadio,S.
  TITLE     Organization of the teicoplanin gene cluster in Actinoplanes
            teichomyceticus
  JOURNAL   Microbiology (Reading, Engl.) 150 (PT 1), 95-102 (2004)
   PUBMED   14702401
REFERENCE   2  (bases 1 to 89713)
  AUTHORS   Donadio,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2003) Donadio S., Microbial Technologies, Vicuron
            Pharmaceuticals, via R. Lepetit 34, 21040 Gerenzano, ITALY
FEATURES             Location/Qualifiers
     source          1. .89713
                     /organism="Actinoplanes teichomyceticus"
                     /mol_type="genomic DNA"
                     /strain="ATCC31131"
                     /db_xref="taxon:1867"
     gene            complement(41. .529)
                     /gene="tei1"
     CDS             complement(41. .529)
                     /gene="tei1"
                     /function="unknown"
                     /note="similar to S. coelicolor hypothetical protein
                     CAB95799.1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAE53333.1"
                     /db_xref="GI:40713133"
                     /db_xref="UniProt/TrEMBL:Q70B16"
                     /translation="MTDAVMEEIRRAQEMNLGGDRAGARATLERLWETADDALHRCTI
                     AHYLADLQGSVEDELRWDELALREFPGLTDERAREYDATWQARAFLPSLHLNLADVHR
                     RAGNRGRARHHLTEAVAALDALPADDYGAMIRDALDEVREAIDTGSTAPLRPASQAQK
                     KP"
     gene            complement(1369. .1899)
                     /gene="tei2"
     CDS             complement(1369. .1899)
                     /gene="tei2"
                     /function="unknown"
                     /note="similar to X. campestris hypothetical protein
                     AAM40941.1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAE53334.1"
                     /db_xref="GI:40713134"
                     /db_xref="InterPro:IPR010556"
                     /db_xref="UniProt/TrEMBL:Q70B15"
                     /translation="MLELITVGHGTAGRDELAALFTGAGLDRVVDVRRYPGSRHHPHV
                     GVEALSRWLPEAGIAYRADQRLGGRRRLPPESPDIWWRVEAFRAYAAHMRTAEFHEAV
                     DELLTDIRAGTTAVMCSETLWWRCHRRLIADFVSEIHGFPVRHLDHRGRLARHRVAAG
                     ARWRSPDELVYDAAPP"
     gene            2346. .3476
                     /gene="tei3"
     CDS             2346. .3476
                     /gene="tei3"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="putative RCC1 repeats protein"
                     /protein_id="CAE53335.1"
                     /db_xref="GI:40713135"
                     /db_xref="InterPro:IPR000408"
                     /db_xref="UniProt/TrEMBL:Q70B14"
                     /translation="MSTGIAVAAVAATAGSASASPALAAAPPATVAAGYAHTCAIRTG
                     GTLWCWGSNANGQLGDGTKTSRTTAVRVGDATTWAGVDAGTSHNCAVRTDGTLWCWGS
                     NSRGQLGDGSTTDRTAPVRVGTATNWTTVSAGDRHTCAVRTDGTLWCWGFNRLGQLGD
                     GAQVYVATTPLQVGTASTWASVTTGFAHTCGTRTDGTLWCWGDSSAGQLGLGILSYRT
                     TPGQVGTATNWATVTAGYTFTCATRTDGTLWCFGENGYGQLGVTGTYQASPVQVGSVA
                     TWTGVAAGFDTACASRADGSLWCWGNNAAGQIGDGTTTHRYAPTRVGTATTWTAALAA
                     GYHNCGFRTDDSLWCWGSNANGELGDGTTTARPAPQQVSLPA"
     gene            complement(3552. .4628)
                     /gene="tei4"
     CDS             complement(3552. .4628)
                     /gene="tei4"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glucanase"
                     /protein_id="CAE53336.1"
                     /db_xref="GI:40713136"
                     /db_xref="GOA:Q70B13"
                     /db_xref="InterPro:IPR001919"
                     /db_xref="InterPro:IPR003961"
                     /db_xref="UniProt/TrEMBL:Q70B13"
                     /translation="MPHRPRRIVTALVLTAALAVAGAVTGPVAHAGDRPATSTPTPPN
                     PSPFPPTAPAGLTAAAVHATSVTLTWTASRPGCCVVDHYEVQYRLAFNDVVALVDAGN
                     VTSTTITGLAPARQYSFQVTAVDGVNHRSAPSTVVTVVTPVSDTAPDTSPPSAPTGLA
                     ATEVTASRAVLTWAPSTDDVGVTGYDVYRFDGWYTSTLLTTTTATTYAVPTGGGRNVY
                     YVRARDAAGNVSIASDTVSTVITTPSPSTPPPPEPVCRVGYRATSTWAGGFVAEVTVT
                     NIGTEPIDGWTLTFAYGGDQRVSASWNGDVTQSGADVTVTGAAWNRRLGAGASATVGV
                     LGSWRTSDAPPAAYTLNGRSCVTA"
     gene            5079. .6608
```

```
                /gene="tei5"
CDS             5079. .6608
                /gene="tei5"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /product="putative membrane zinc protease"
                /protein_id="CAE53337.1"
                /db_xref="GI:40713137"
                /db_xref="GOA:Q70B12"
                /db_xref="InterPro:IPR001570"
                /db_xref="InterPro:IPR005075"
                /db_xref="InterPro:IPR006025"
                /db_xref="UniProt/TrEMBL:Q70B12"
                /translation="MVRTPLAVALAAGTVAASALVAVSGPAAAQAPAPSVDAVAQADA
                VVAARPAVLQAGKHDDFHRRQVYKSQGLTYASYDRTYKGFPVKGGDFVVVTDNAGQTR
                YTSVAQSSPIGELSTEPTVSESAAIATAKKQLKSVTSVEATRLVVVAAEGRAAKLAYE
                TTVNGTGAEGVSRLTVDVDATTGAVLATEEHVTEATGTGTGWIVGSVSLNTTQSGSTY
                SLKDPTVTNLSCQDASTRTTFSGSDNVWGNGTGTSKETGCVDALYAAQKQTAMLSAWL
                GRNGQNGSGGAWPIRVGLNQQNAYYDGTQIQIGKNTAGKWISSLDVVGHELGHGIDDK
                TPGGISKGGTQEFVADTIGAATEAYAANPNDPFDYQVGEEVNLVGSGPIRYMYNPSLA
                GDDNCYSSSTPTDEVHAAAGPGNHWFYLLAQGSGGNGQPASSTCNGSTVTGVGIQTAI
                KIMYNAMLLKTSSSSYAKYRVWTLQAAKNLYPGSCTQFNAVKAAWNAVSLPAQSGEPT
                C"
gene            complement(6744. .7079)
                /gene="tei6"
CDS             complement(6744. .7079)
                /gene="tei6"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /product="putative mannose-6-phosphate isomerase"
                /protein_id="CAE53338.1"
                /db_xref="GI:40713138"
                /db_xref="GOA:Q70B11"
                /db_xref="InterPro:IPR007113"
                /db_xref="UniProt/TrEMBL:Q70B11"
                /translation="MLTKITPAAEAASLDEYWSQRVLTAANGNLFKVAKGLGSTRWHA
                HDDQDETFLMLAGTLTVHLRDGSVELTPGDLLVIPRGVEHRPEAAGEARFLLVGPEVT
                STEAGGKPA"
gene            complement(7158. .8123)
                /gene="tei7"
CDS             complement(7158. .8123)
                /gene="tei7"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /product="putative AraC-family transcription regulator"
                /protein_id="CAE53339.1"
                /db_xref="GI:40713139"
                /db_xref="GOA:Q6ZZK7"
                /db_xref="InterPro:IPR000005"
                /db_xref="InterPro:IPR002818"
                /db_xref="UniProt/TrEMBL:Q6ZZK7"
                /translation="MVPRRVALIVDEHSNPFEVGCACEIFGGRRRPQIGYELYELTVV
                SPTRSVPMRDRLFRVSAPGRLSDLDRADMIIVPNRPDVDVPSRGAVLAAIRRAHARGA
                RLVGLCTGAFTLAEAGVLDGRRAAVHWQLAPDFHRRFPAVRLEPDVLFVDDGDILTSA
                GSAAALDLGLHIVRRDHGTEVANRVGRQLVFAAFRDGGQRQFVERPLPAPATSPLAPV
                LQWAQERLHLRLSVDDLARHGRMSTATLHRRFRNEVGSTPLAWLTAQRVALACRLIEQ
                GAAGVDAVARQSGLGTAANLRALLRKHVGLSPSEYRRRFAADGAR"
gene            complement(8427. .9179)
                /gene="tei8"
CDS             complement(8427. .9179)
                /gene="tei8"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /product="putative TetR-family transcitpional regulator"
                /protein_id="CAE53340.1"
                /db_xref="GI:40713140"
                /db_xref="GOA:Q6ZZK6"
                /db_xref="InterPro:IPR001647"
                /db_xref="UniProt/TrEMBL:Q6ZZK6"
```

```
  Query Match            23.9%;  Score 70.4;  DB 1;  Length 89713;
  Best Local Similarity  54.7%;  Pred. No. 1.1e+02;
  Matches  162;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    1;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              | |  | |||  ||      | || |||| ||       ||||||  |    ||| ||||
Db      45852 GTGCACCGTCACCGACAGCACGTGCTCCTTCTCCGACAGCGCGAAGAGCTCGCACCGCCA 45793

Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCG---CGGCCC 119
              || |  | |   |  | || ||  |||  || ||     || ||||   |||    ||
Db      45792 CGGCATGTCCCGGGTGAGGTCGAACAGCTGCTCGCCCCGCTCGGCCAGCCGTGCGGGGAG 45733

Qy        120 CCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAG 179
                |  ||||  |||   || | ||| | | || ||  || ||  | ||||   |||  |
Db      45732 GTCCTCCTCGGTGGCCGGGACCGGCGTCAGCCGGACCGGCGCGGCGTCGTGGATGTGCTG 45673

Qy        180 CGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGAT 239
                ||| | |||||||  | |||  ||  | | ||  |    ||  | || |  |||
Db      45672 GTGGACGGTCTGCGCGTCGCCCGGGAAGGTGGTCCGCAGGATCTCGTGCCGGGCCGCGAC 45613

Qy        240 GGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |  || || |  ||| |||||| | |   ||||  |  ||||| ||||| ||||||
Db      45612 GTCGCGCAGCGCCGCCTCCAGGGCGGCCACGTCCAGCCGGCCGCGCAGGCGCAGCG 45557
```

```
RESULT 6
AJ632270/c
LOCUS       AJ632270               89976 bp    DNA     linear   BCT 15-APR-2005
DEFINITION  Actinoplanes teichomyceticus teicoplanin gene cluster.
ACCESSION   AJ632270
VERSION     AJ632270.1  GI:45580837
KEYWORDS    ABC transporter; Acyl-CoA synthase; acyltransferase; b-hydroxylase;
            D-lactate dehydrogenase; dehydrogenase; dihydrofolate reductase;
            dipeptidase; DpgA protein; DpgB protein; DpgC protein; DpgD
            protein; glycosyltransferase; GTP cyclohydrolase I; halogenase;
            heptulosonate synthase; HmaS protein; HmO protein; HpgT protein;
            ligase; LuxR regulatory protein; manganese ABC transporter;
            mannossyltransferase; MurF-like protein; Na+-H+ antiporter; nonheme
            iron dioxygenase; P450 monooxygenase; peptide synthetase;
            prephenate dehydrogenase; response regulator; teiA gene; teiB gene;
            teiC gene; teiD gene; thioesterase; transcriptional regulator;
            transposase.
SOURCE      Actinoplanes teichomyceticus
  ORGANISM  Actinoplanes teichomyceticus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Micromonosporineae; Micromonosporaceae; Actinoplanes.
REFERENCE   1
  AUTHORS   Li,T.L., Huang,F., Haydock,S.F., Mironenko,T., Leadlay,P.F. and
            Spencer,J.B.
  TITLE     Biosynthetic gene cluster of the glycopeptide antibiotic
            teicoplanin: characterization of two glycosyltransferases and the
            key acyltransferase
  JOURNAL   Chem. Biol. 11 (1), 107-119 (2004)
   PUBMED   15113000
REFERENCE   2  (bases 1 to 89976)
  AUTHORS   Li,T.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2004) Li T.L., Department of Chemistry,
            University of Cambridge, Lensfield Road, CB2 1EW, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .89976
                     /organism="Actinoplanes teichomyceticus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1867"
     CDS             complement(552. .1517)
                     /note="ORF11"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator"
                     /protein_id="CAG14998.1"
                     /db_xref="GI:45580838"
                     /db_xref="GOA:Q6ZZK7"
```

```
                     /db_xref="InterPro:IPR000005"
                     /db_xref="InterPro:IPR002818"
                     /db_xref="UniProt/TrEMBL:Q6ZZK7"
                     /translation="MVPRRVALIVDEHSNPFEVGCACEIFGGRRRPQIGYELYELTVV
                     SPTRSVPMRDRLFRVSAPGRLSDLDRADMIIVPNRPDVDVPSRGAVLAAIRRAHARGA
                     RLVGLCTGAFTLAEAGVLDGRRAAVHWQLAPDFHRRFPAVRLEPDVLFVDDGDILTSA
                     GSAAALDLGLHIVRRDHGTEVANRVGRQLVFAAFRDGGQRQFVERPLPAPATSPLAPV
                     LQWAQERLHLRLSVDDLARHGRMSTATLHRRFRNEVGSTPLAWLTAQRVALACRLIEQ
                     GAAGVDAVARQSGLGTAANLRALLRKHVGLSPSEYRRRFAADGAR"
CDS                  complement(1820. .2572)
                     /note="ORF10"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator"
                     /protein_id="CAG14999.1"
                     /db_xref="GI:45580839"
                     /db_xref="GOA:Q6ZZK6"
                     /db_xref="InterPro:IPR001647"
                     /db_xref="UniProt/TrEMBL:Q6ZZK6"
                     /translation="MSATQEPGTSSRRRGGAGPVDDATVRAALIQAAERQLTASPDGD
                     IATRAVSEEAGVTQPVLYRLFGDKRGLLDAVADAGFDRYARRKSELEVTDNPVGDLYA
                     GWDDHMAFATENPALYQLMFAPRPRSGATTYRRILTLLEATLLRCAAAGALTTTPRAA
                     AQLILSANIGVALGRIAAPELFDDPALSHRARDAAFATVLTHPATPSGPDPLRDAARQ
                     LRSQLSLSGTDRLEPAETVLLDRWLERIDQPE"
CDS                  2684. .3418
                     /note="ORF9"
                     /codon_start=1
                     /transl_table=11
                     /product="short chain dehydrogenase"
                     /protein_id="CAG15000.1"
                     /db_xref="GI:45580840"
                     /db_xref="GOA:Q6ZZK5"
                     /db_xref="InterPro:IPR002198"
                     /db_xref="InterPro:IPR002347"
                     /db_xref="UniProt/TrEMBL:Q6ZZK5"
                     /translation="MGDHRVAVVTGASRGIGRAVATALAGAGHAVVVGYAGNEAAAKE
                     VVDRIAATGGRAVAARADVADENAVAALFDTATEAFGGVDVVVNSAGRMALSPIADLD
                     LADLDALHRTNIRGTFVVAREAARRVRPGGAIVLVSTSILGLQFPGYGAYAASKGAVE
                     AMTLILAREMRGRDITVNTVAPGPTATELFLEGKDEATVEHLAKQPPLERLGTPADIA
                     GVVAFLAGPGGRWVNGQVLRANGGIV"
CDS                  4070. .5413
                     /function="addition of D-Ala"
                     /note="ORF8"
                     /codon_start=1
                     /transl_table=11
                     /product="MurF-like protein"
                     /protein_id="CAG15001.1"
                     /db_xref="GI:45580841"
                     /db_xref="GOA:Q6ZZK4"
                     /db_xref="InterPro:IPR000713"
                     /db_xref="InterPro:IPR004101"
                     /db_xref="InterPro:IPR005863"
                     /db_xref="UniProt/TrEMBL:Q6ZZK4"
                     /translation="MIPLSLEAIAAVVGGTVEGDHTVTVTAPAVLDGRQAEPGGLFVA
                     FAGEHVDGHDYAGQAGQAGAVAVLGSRPTALPTVVVEDARTALQALAADVVARLRERL
                     TLVGVTGSQGKTSTKDLLAAVLSSAAPTIATKGNLNNELGVPLTMLRAGAATRFLVVE
                     MGARRIGDLAKLTDLVPLDVAVVLNVGEAHLGVFGSRAAIAEGKGELVRGLVPGGTAI
                     LNADDPRVVAMRSLTDGPVLTFGHAEDADVRVRDLTLDRLGRPSFTLRTADATARVKL
                     PIVGAHQALNAAAAVAAALSAGLTLDAAAATLATASLTRWRLELRDIAGGVTLLNDSY
                     NASPGSTRAALDALAAIEGGRRVAVLGQMLELGDESEAEHRAVGEYAAARADVVVTVG
                     EAVRPLAVGAGERAVAVADNGAAVDWLRGHLAPGDVVLVKASRGARLDEVAAALS"
CDS                  5612. .6652
                     /note="ORF7"
                     /codon_start=1
                     /transl_table=11
                     /product="D-lactate dehydrogenase"
                     /protein_id="CAG15002.1"
                     /db_xref="GI:45580842"
                     /db_xref="GOA:Q6ZZK3"
                     /db_xref="InterPro:IPR006139"
                     /db_xref="InterPro:IPR006140"
                     /db_xref="UniProt/TrEMBL:Q6ZZK3"
                     /translation="MGYSEPVRAAAPRTRSLASSSSPVVPATGITIYGCGPDEAALFR
                     ELAPRFGVLPIITEAAVSEANVNLASGNRCISVGHKTQITNGTLLALSRSGVKYISTR
                     SIGYNHIDVAYAQSVGISVGNVAYSPDSVADYTLMLMLMAVRHMKAMIRRADVHDYRL
                     NDVRGKELRDLTVGVVGTGRIGSAVMDRLWGFGCRVLAYDNRRSDSAHYVPLDDLVRQ
                     SDIVTLHTPLTAATHHLLNQDRIAQMRNGAFVINTGRGSLIDTEALVAALESDKLGGA
                     ALDVLEGEEGVFYADCRNKLIESKALLRLQEMPNVIISPHMAYYTDHALSDTVENSLV
                     NCMNFEREQQHG"
CDS                  6645. .7682
                     /function="D-Ala- D-Lac ligase"
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="ligase"
                     /protein_id="CAG15003.1"
                     /db_xref="GI:45580843"
                     /db_xref="GOA:Q6ZZK2"
                     /db_xref="InterPro:IPR000291"
                     /db_xref="InterPro:IPR005905"
                     /db_xref="InterPro:IPR011095"
                     /db_xref="InterPro:IPR011127"
                     /db_xref="InterPro:IPR011761"
                     /db_xref="UniProt/TrEMBL:Q6ZZK2"
                     /translation="MGKLKIGIIFGGVSEEHPVSIKSAQEVAKSLDAEKYEPYWIGIT
                     KDGAWRLCDGPGADWEISSARPAMLSPDRNVHGLLVQEDGRYREIRLDVVLPVLHGKL
                     GEDGAMQGLFELSGIPYVGCDVQSSAVCMDKSLTYAVVSSAGIAAATFWVVTPDEAPD
                     PAQLTYPVFVKPARSGSSFGVTKVSTEQELKAALETAREFDSKVLIEEAVVGSEIGCA
                     IMGNDPDLVAGEVDRVALSHGFFRIHQESEPERGSENSTFICPADIPAEARSLVQETA
                     KAIYRALGCRGIARVDMFLKEDGTVVLNEVNTLPGLTSYSRYPRMMAAAGIPLSAVID
                     QIVSLALPGDI"
CDS                  7682. .8290
                     /function="D-Ala- D-Ala dipeptidase"
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="dipeptidase"
                     /protein_id="CAG15004.1"
                     /db_xref="GI:45580844"
                     /db_xref="GOA:Q6ZZK1"
                     /db_xref="InterPro:IPR000755"
                     /db_xref="UniProt/TrEMBL:Q6ZZK1"
                     /translation="MKDDFAFVDELVPGIRWDAKYATWDNFTGKPVDGYVANRIVGTK
                     ALCAALARAQEEAETLGFGLLLWDGYRPQRAVDCFMRWAEQPEDGRKKLRHYPNIGRP
                     EMFEKGYVATKSGHSRGSTVDLTLYNLATGELASMGGDHDLMDPISHHGASGITPIEA
                     KNREYLCSIMEGCGFARYDYEWWHYTLKNEPYPDTYFDFPIV"
CDS                  complement(9052. .9591)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydrofolate reductase"
                     /protein_id="CAG15005.1"
                     /db_xref="GI:45580845"
                     /db_xref="GOA:Q6ZZK0"
                     /db_xref="InterPro:IPR002734"
                     /db_xref="UniProt/TrEMBL:Q6ZZK0"
                     /translation="MSDVSGAGKVVVNRAMSLDGFIAGPGHAMDWIFEYLTGDMFPEV
                     MASTGAMLIGRGTYEVAKRMAGQGTPYDGGAQFVLTHEPPHEPDPAVTFLSCDIEEAV
                     ATARSAAGGKNLEILGADVAAQCLRRGLVDEILVYVVPVLLGDGVRFSPPSLDRIDLE
                     PLSNTRSGAVTMLRFRVRK"
CDS                  complement(9704. .10795)
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator"
                     /protein_id="CAG15006.1"
                     /db_xref="GI:45580846"
                     /db_xref="GOA:Q6ZZJ9"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="InterPro:IPR003660"
                     /db_xref="InterPro:IPR003661"
                     /db_xref="InterPro:IPR004358"
                     /db_xref="InterPro:IPR005467"
                     /db_xref="UniProt/TrEMBL:Q6ZZJ9"
                     /translation="MDRAPGLSVRLKLTLSYAGFLMLAGASLLATVWVYLLRGLPDTA
                     MSSRGLVPIKTIVLHAFAPATAGVLLFLLVFGLLGGWLLAGRMLAPLTRITEATRMAA
```

```
                NGSLSHRIELDGRKDEFRELADAFDIMLTRLEAHIAEHQRFAANASHELRTPLAITKT
                LLDVARTDPARSTHELIGRLHTVNTRAIDLTEALLLLSRADRRSFKREYVDLSLNVEE
                AIETLLPLAEKHGVTIEASGDITPTVGSRALLLQMTINLLHNAIVHNVAEERIVWVTT
                SVHPKSVTLTVENTGEQLSAQLVSTLTEPFQRGTERIHHDHAGVGLGLAIVKSIAHAH
                DGTLTTAPRAGGGLHVTVQLPAATTPPRN"
     CDS        complement(10803. .11483)
                /note="ORF2"
                /codon_start=1
                /transl_table=11

  Query Match             23.9%;  Score 70.4;  DB 1;  Length 89976;
  Best Local Similarity   54.7%;  Pred. No. 1.1e+02;
  Matches  162;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    1;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              |  |  | |||  ||     | || |||| ||      ||||||  |    ||| ||||
Db      39245 GTGCACCGTCACCGACAGCACGTGCTCCTTCTCCGACAGCGCGAAGAGCTCGCACCGCCA 39186

Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCG---CGGCCC 119
              || |  | |   |  | || ||  |||  || ||    || ||||   |||    ||
Db      39185 CGGCATGTCCCGGGTGAGGTCGAACAGCTGCTCGCCCCGCTCGGCCAGCCGTGCGGGGAG 39126

Qy        120 CCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAG 179
                |  ||||  |||   || | ||| | | || ||  || ||  | ||||   |||  |
Db      39125 GTCCTCCTCGGTGGCCGGGACCGGCGTCAGCCGGACCGGCGCGGCGTCGTGGATGTGCTG 39066

Qy        180 CGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGAT 239
                 ||| | |||||||  | |||  ||  | | ||  |    ||  | || | |||
Db      39065 GTGGACGGTCTGCGCGTCGCCCGGGAAGGTGGTCCGCAGGATCTCGTGCCGGGCCGCGAC 39006

Qy        240 GGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |  || || |  ||| |||||| | |   ||||  |  ||||| ||||| ||||||
Db      39005 GTCGCGCAGCGCCGCCTCCAGGGCGGCCACGTCCAGCCGGCCGCGCAGGCGCAGCG 38950


RESULT 7
AC143286
LOCUS       AC143286               167624 bp    DNA     linear   HTG 09-APR-2003
DEFINITION  Macaca mulatta clone CH250-270J2, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC143286
VERSION     AC143286.1  GI:29567925
KEYWORDS    HTG; HTGS_PHASE2; HTGS_PGI.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 167624)
  AUTHORS   Csuros,M. and Milosavljevic,A.
  TITLE     Pooled genomic indexing (PGI): mathematical analysis and experiment
            design
  JOURNAL   (in) Guigo,R. and Gusfield,D. (Eds.);
            ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
            2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;
            Springer (2002)
REFERENCE   2  (bases 1 to 167624)
  AUTHORS   Milosavljevic,A., Sodergren,E., Csuros,M., Li,B., Jackson,A.R.,
            Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L.,
            Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Egan,A., Earnhart,C., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 167624)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 167624)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: LAKZ
                Center clone name: CH250-270J2
            -------------- Summary Statistics
                Chemistry: Dye-primer Bodipy: inf% of reads
                Chemistry: Dye-terminator Big Dye: inf% of reads
                Consensus quality: 6374 bases at least Q40
                Consensus quality: 8322 bases at least Q30
                Consensus quality: 10272 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: The contigs are based on the application
            * of the PGI method using the Human genome (NCBI build 31)
            * as the comparative genome.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   167624: contig of 167624 bp in length.
FEATURES            Location/Qualifiers
     source         1. .167624
                    /organism="Macaca mulatta"
                    /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9544"
                     /clone="CH250-270J2"
     misc_feature    1. .167624
                     /note="assembly_name:CH250-270J2.1A
                     CONFIDENCE:_0.83"
ORIGIN

  Query Match             23.7%;  Score 70;  DB 14;  Length 167624;
  Best Local Similarity   54.3%;  Pred. No. 96;
  Matches  159;  Conservative    0;  Mismatches  130;  Indels    4;  Gaps    1;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              |     ||   | | |  | ||    | |  ||  || ||||||||||| ||| || | |
Db      54418 GNNNNNGGGGGGGGGGGGGGGGGGGGCGGGGCGCGCGGGCGCGCGGGCGCCGCCCCCCGC 54477

Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCT 122
               |||   | |||||||| | || || | ||  || |  ||| |||  | ||||  |
Db      54478 GCGGGCGGGCCCCCCGCGGGGGCGCGGGGGGGCCGGGCGGCCGGGGGCGGGGGGCCGGCC 54537

Qy        123 CCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGC 182
               ||| || | | | |    |    || ||||| ||||  |||||  |  || | ||  |||
Db      54538 GCGCGCGGGCGGGGCCGGC----CCGGGCGCCCGCGGGCGCGGCGGGCCGGGCGGGCCGC 54593

Qy        183 CTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGG 242
               | |  ||||  | |||||| ||| || || | || || |    ||| |  |  |||
Db      54594 GCGCGCCCCCGCGGGCGCGGGGGGGGGGGGGGGGGGGGGGCGGCGGGGGCGGGGGGCGGG 54653

Qy        243 GCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |    | | ||  |  || | | | | | | |||  | |  || |  | |
Db      54654 GGGGGCGGGGGCGGCGGGGGCGCGGCGGGGGCGCGGGGGGGGGCGGGGGGGG 54706


RESULT 8
AC092720
LOCUS       AC092720               169214 bp    DNA     linear   PRI 18-DEC-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-482M8, complete sequence.
ACCESSION   AC092720 AC012002
VERSION     AC092720.4  GI:40018661
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 169214)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169214)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 169214)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 169214)
  AUTHORS   Stanford Human Genome Center and Los Alamos National Laboratory.
  CONSRTM   DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 18, 2003 this sequence version replaced gi:18542979.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.5.

FEATURES             Location/Qualifiers
     source          1. .169214
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-482M8"
ORIGIN

  Query Match             23.4%;  Score 69;  DB 8;  Length 169214;
  Best Local Similarity   55.7%;  Pred. No. 1.2e+02;
  Matches  132;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         45 GCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGC 104
              || | | ||||  |||  || ||||| | ||  |   ||| |  ||||  || || ||||
Db      22664 GCCGCCCGCGCCGCGCCGCGACCCGGGGCCGGGGACCGCGCGGGGGCGGCCGGGGGGCGC 22723

Qy        105 GGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCG 164
               | || ||||||||     |    || |  ||  ||   |||  | | |  || |  || |
Db      22724 CGGGGGCCGCGGCCGGGGCGCGCCGGGACCGGGGACGCGCGGGCCGGGGCCGGGTCCGGG 22783

Qy        165 GGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCC 224
               | | ||||| ||    |  || | |   | | |||||  |||  | ||  ||| |||
Db      22784 CGAGGGCGCGCGGGCGCCGCGGGCCGGGCGGGCGCCGGGGCGGGGGCGGGGACCGGGGCC 22843

Qy        225 GGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCC 281
              || || | |  |  |  ||  |    | ||   |  |  |||  || || |||||||
Db      22844 GGGCAAGGCTCCCGTCAGGGGCCGGCGGGCGGGCGCGCGCGGGGCGCCGGGGGGGCC 22900


RESULT 9
PM2H12G
LOCUS       PM2H12G                1279 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm2h12.g, sequence tagged site.
ACCESSION   AL684840
VERSION     AL684840.1  GI:19337636
KEYWORDS    STS.
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 1279)
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .1279
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm2h12.g"
ORIGIN

  Query Match             23.0%;  Score 67.8;  DB 10;  Length 1279;
  Best Local Similarity   51.0%;  Pred. No. 9.3e+02;
  Matches  146;  Conservative    0;  Mismatches  139;  Indels    1;  Gaps    1;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||  ||  |  | ||||  || |      ||| |    |  | || || ||||||||||||
Db        713 GCCGGGNNGCGCNGCGCCCCGCGCCGGGCCGGCGCCCGCGGGGCCCGCGCGGCGCGCCGC 772

Qy         61 -TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCC 119
                ||  |||||| ||||  ||| |  | || |   | |  ||||||  | || || | |
Db        773 GGNCGCGCCGGCGCCGCCNNGGGNGGCCGGGGGGCGGCGCCGCGCGGGCGGCCCCGCCGC 832

Qy        120 CCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAG 179
```

```
                 |  ||| || ||  | ||     ||| || ||||||||   |   |||    ||  |
Db        833 GCCNCGCGCGCGGCGGGCGCGGGCCGCGCGCCGCACGCGNGGGGNNGCGCCGCCGGCGCC 892

Qy        180 CGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGAT 239
              |||  | | | |  || ||    || | | |   |    | | | |    ||||  |
Db        893 CGCGCCCGGCGGGCCGCCCGNNCGGCCCGCCCNNGCNCGGCCCCGCGNNGCGCCGCGGCG 952

Qy        240 GGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGC 285
                | |   |  | |  ||  |  ||    | || | ||  |  ||
Db        953 NNNGNCNCGCCCCCCGGCCNCGGGCGNNGGGCCGCGNGGCGCCGGC 998


RESULT 10
AC139773/c
LOCUS       AC139773               65351 bp    DNA     linear   HTG 13-FEB-2003
DEFINITION  Homo sapiens chromosome 19 clone LLNLF-200C8, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC139773
VERSION     AC139773.1  GI:28372600
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 65351)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65351)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 19502, F24136
            Center clone name: LLNL-F_200C8
            -------
            Summary Statistics
            Consensus quality: 44919 bases at least Q40
            Consensus quality: 48008 bases at least Q30
            Consensus quality: 51090 bases at least Q20
            Estimated insert size: 37980; agarose-fp estimation
            Estimated insert size: 63251; sum-of-contigs estimation
            Quality coverage:  37.55 in Q20 bases; agarose-fp estimation
            Quality coverage:  22.55 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1361: contig of 1361 bp in length
            *     1362     1461: gap of unknown length
            *     1462     2597: contig of 1136 bp in length
            *     2598     2697: gap of unknown length
            *     2698     3861: contig of 1164 bp in length
            *     3862     3961: gap of unknown length
            *     3962     5028: contig of 1067 bp in length
            *     5029     5128: gap of unknown length
            *     5129     6316: contig of 1188 bp in length
            *     6317     6416: gap of unknown length
            *     6417     7456: contig of 1040 bp in length
            *     7457     7556: gap of unknown length
            *     7557     8799: contig of 1243 bp in length
            *     8800     8899: gap of unknown length
            *     8900    10206: contig of 1307 bp in length
            *    10207    10306: gap of unknown length
            *    10307    11360: contig of 1054 bp in length
            *    11361    11460: gap of unknown length
            *    11461    12575: contig of 1115 bp in length
            *    12576    12675: gap of unknown length
            *    12676    14195: contig of 1520 bp in length
            *    14196    14295: gap of unknown length
            *    14296    15537: contig of 1242 bp in length
            *    15538    15637: gap of unknown length
            *    15638    16715: contig of 1078 bp in length
            *    16716    16815: gap of unknown length
            *    16816    17986: contig of 1171 bp in length
            *    17987    18086: gap of unknown length
            *    18087    19591: contig of 1505 bp in length
            *    19592    19691: gap of unknown length
            *    19692    20898: contig of 1207 bp in length
            *    20899    20998: gap of unknown length
            *    20999    22173: contig of 1175 bp in length
            *    22174    22273: gap of unknown length
            *    22274    23837: contig of 1564 bp in length
            *    23838    23937: gap of unknown length
            *    23938    24994: contig of 1057 bp in length
            *    24995    25094: gap of unknown length
            *    25095    26961: contig of 1867 bp in length
            *    26962    27061: gap of unknown length
            *    27062    33488: contig of 6427 bp in length
            *    33489    33588: gap of unknown length
            *    33589    65351: contig of 31763 bp in length.
FEATURES             Location/Qualifiers
     source          1. .65351
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="LLNLF-200C8"
                     /clone_lib="Lawrence Livermore human cosmid library LLNLF"
     gap             1362. .1461
                     /estimated_length=unknown
     gap             2598. .2697
                     /estimated_length=unknown
     gap             3862. .3961
                     /estimated_length=unknown
     gap             5029. .5128
                     /estimated_length=unknown
     gap             6317. .6416
                     /estimated_length=unknown
     gap             7457. .7556
                     /estimated_length=unknown
     gap             8800. .8899
                     /estimated_length=unknown
     gap             10207. .10306
                     /estimated_length=unknown
     gap             11361. .11460
                     /estimated_length=unknown
     gap             12576. .12675
                     /estimated_length=unknown
     gap             14196. .14295
                     /estimated_length=unknown
     gap             15538. .15637
                     /estimated_length=unknown
     gap             16716. .16815
                     /estimated_length=unknown
     gap             17987. .18086
                     /estimated_length=unknown
     gap             19592. .19691
                     /estimated_length=unknown
     gap             20899. .20998
                     /estimated_length=unknown
     gap             22174. .22273
                     /estimated_length=unknown
     gap             23838. .23937
                     /estimated_length=unknown
```

```
     gap             24995. .25094
                     /estimated_length=unknown
     gap             26962. .27061
                     /estimated_length=unknown
     gap             33489. .33588
                     /estimated_length=unknown
ORIGIN

  Query Match            23.0%;  Score 67.8;  DB 14;  Length 65351;
  Best Local Similarity  52.7%;  Pred. No. 2.1e+02;
  Matches  147;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy          7 GCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTC 66
              |||| | | ||| |  |  ||  ||| |||||| ||| || |    ||| || |   | |
Db       2447 GCGGCCCGGGCGCCCCGCCCCGCCGGGCGGCCGGGCGGGCCGCGCCCGCCCCCCGCGGGC 2388

Qy         67 CCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGC 126
                |||  | |  |  |    |  |||||  || |    ||||  | ||  ||||| |||
Db       2387 GGCGCGCGGGCCGGCCCCCCCGCCCGCGCCCGGGGGCGCGCGGGCGGCCCCCCCCCCGG 2328

Qy        127 TCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCG 186
                | | ||      | || ||| || | | | | |  |||||   |    || || || ||
Db       2327 GGGGGCGCCGGCCCCCGCCCCCCGCCCCCCCCCCCCGGGGCGCCCCGCCGGGGCCCCGCG 2268

Qy        187 GACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCC 246
              | |    | |||||    | | |||||||| |||  |   | |  |  ||||  |
Db       2267 CCGCGGGGCCGCCGGGGGGGCCCCCGGGGCGCGGCCCCGCCCCCCCGGGGGGCGGGGGGC 2208

Qy        247 ACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGC 285
              || |||   |  |   | |   |  | |||| ||  ||
Db       2207 GCCGGCGGCCCGGGCCGCCGGGGGGGGGGGGGCCCCGGC 2169


RESULT 11
AC026284/c
LOCUS       AC026284              156842 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-536F13, WORKING DRAFT
            SEQUENCE, 30 unordered pieces.
ACCESSION   AC026284
VERSION     AC026284.3  GI:8570286
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 156842)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156842)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 17, 2000 this sequence version replaced gi:7344515.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0536F13
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 141875 bases at least Q40
            Consensus quality: 145977 bases at least Q30
            Consensus quality: 148333 bases at least Q20
            Insert size: 152000; agarose-fp
            Insert size: 153942; sum-of-contigs
            Quality coverage: 3.83 in Q20 bases; agarose-fp
            Quality coverage: 3.88 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1157: contig of 1157 bp in length
            *     1158     1257: gap of unknown length
            *     1258     2298: contig of 1041 bp in length
            *     2299     2398: gap of unknown length
            *     2399     3770: contig of 1372 bp in length
            *     3771     3870: gap of unknown length
            *     3871     5999: contig of 2129 bp in length
            *     6000     6099: gap of unknown length
            *     6100     8944: contig of 2845 bp in length
            *     8945     9044: gap of unknown length
            *     9045    11261: contig of 2217 bp in length
            *    11262    11361: gap of unknown length
            *    11362    13705: contig of 2344 bp in length
            *    13706    13805: gap of unknown length
            *    13806    15970: contig of 2165 bp in length
            *    15971    16070: gap of unknown length
            *    16071    18910: contig of 2840 bp in length
            *    18911    19010: gap of unknown length
            *    19011    22444: contig of 3434 bp in length
            *    22445    22544: gap of unknown length
            *    22545    25041: contig of 2497 bp in length
            *    25042    25141: gap of unknown length
            *    25142    27895: contig of 2754 bp in length
            *    27896    27995: gap of unknown length
            *    27996    31827: contig of 3832 bp in length
            *    31828    31927: gap of unknown length
            *    31928    33731: contig of 1804 bp in length
            *    33732    33831: gap of unknown length
            *    33832    37723: contig of 3892 bp in length
            *    37724    37823: gap of unknown length
            *    37824    40576: contig of 2753 bp in length
            *    40577    40676: gap of unknown length
            *    40677    44815: contig of 4139 bp in length
            *    44816    44915: gap of unknown length
            *    44916    47361: contig of 2446 bp in length
            *    47362    47461: gap of unknown length
            *    47462    53565: contig of 6104 bp in length
            *    53566    53665: gap of unknown length
            *    53666    59109: contig of 5444 bp in length
            *    59110    59209: gap of unknown length
            *    59210    65684: contig of 6475 bp in length
            *    65685    65784: gap of unknown length
            *    65785    71929: contig of 6145 bp in length
            *    71930    72029: gap of unknown length
            *    72030    79140: contig of 7111 bp in length
            *    79141    79240: gap of unknown length
            *    79241    86612: contig of 7372 bp in length
            *    86613    86712: gap of unknown length
            *    86713    94446: contig of 7734 bp in length
            *    94447    94546: gap of unknown length
            *    94547   108964: contig of 14418 bp in length
            *   108965   109064: gap of unknown length
            *   109065   122219: contig of 13155 bp in length
            *   122220   122319: gap of unknown length
            *   122320   133948: contig of 11629 bp in length
            *   133949   134048: gap of unknown length
            *   134049   144976: contig of 10928 bp in length
            *   144977   145076: gap of unknown length
            *   145077   156842: contig of 11766 bp in length.
FEATURES             Location/Qualifiers
```

```
source          1. .156842
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="3"
                /clone="RP11-536F13"
misc_feature    1. .1157
                /note="assembly_name:Contig16"
gap             1158. .1257
                /estimated_length=unknown
misc_feature    1258. .2298
                /note="assembly_name:Contig18"
gap             2299. .2398
                /estimated_length=unknown
misc_feature    2399. .3770
                /note="assembly_name:Contig19
                clone_end:T7
                vector_side:right"
gap             3771. .3870
                /estimated_length=unknown
misc_feature    3871. .5999
                /note="assembly_name:Contig20"
gap             6000. .6099
                /estimated_length=unknown
misc_feature    6100. .8944
                /note="assembly_name:Contig21"
gap             8945. .9044
                /estimated_length=unknown
misc_feature    9045. .11261
                /note="assembly_name:Contig22"
gap             11262. .11361
                /estimated_length=unknown
misc_feature    11362. .13705
                /note="assembly_name:Contig23"
gap             13706. .13805
                /estimated_length=unknown
misc_feature    13806. .15970
                /note="assembly_name:Contig24
                clone_end:SP6
                vector_side:left"
gap             15971. .16070
                /estimated_length=unknown
misc_feature    16071. .18910
                /note="assembly_name:Contig25"
gap             18911. .19010
                /estimated_length=unknown
misc_feature    19011. .22444
                /note="assembly_name:Contig26"
gap             22445. .22544
                /estimated_length=unknown
misc_feature    22545. .25041
                /note="assembly_name:Contig27"
gap             25042. .25141
                /estimated_length=unknown
misc_feature    25142. .27895
                /note="assembly_name:Contig28"
gap             27896. .27995
                /estimated_length=unknown
misc_feature    27996. .31827
                /note="assembly_name:Contig29"
gap             31828. .31927
                /estimated_length=unknown
misc_feature    31928. .33731
                /note="assembly_name:Contig30"
gap             33732. .33831
                /estimated_length=unknown
misc_feature    33832. .37723
                /note="assembly_name:Contig31"
gap             37724. .37823
                /estimated_length=unknown
misc_feature    37824. .40576
                /note="assembly_name:Contig32"
gap             40577. .40676
                /estimated_length=unknown
misc_feature    40677. .44815
                /note="assembly_name:Contig33"
gap             44816. .44915
                /estimated_length=unknown
misc_feature    44916. .47361
                /note="assembly_name:Contig34"
gap             47362. .47461
                /estimated_length=unknown
misc_feature    47462. .53565
                /note="assembly_name:Contig35"
gap             53566. .53665
                /estimated_length=unknown
misc_feature    53666. .59109
                /note="assembly_name:Contig36"
gap             59110. .59209
```

```
  Query Match             23.0%;  Score 67.8;  DB 14;  Length 156842;
  Best Local Similarity   54.6%;  Pred. No. 1.5e+02;
  Matches  161;  Conservative    0;  Mismatches  127;  Indels    7;  Gaps    1;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||| ||||      ||| || |   |||| ||||  | ||| | |||| | ||||| |
Db      13669 GCGGCGGCGCGTCAGGCGGCGGGCGGCTCCCCCCGGGGGGGCGGGGGGGCCGGGCGCCCC 13610

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                || | |||||  || | | |||||  ||  || | ||   |||| |    |  |
Db      13609 CGCGCCGCGGCGGTGCGGTGCGCGTGTGGGGTCGTGCGGGTGGCGGGGTGGGGGCGGGGG 13550

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
                |||  |  |  | || ||  | || | ||| |        | ||||||||   || |
Db      13549 GGGCGCGGGGCGCGGGCGCTCGCCCCCGGCCGCCC-------CTGGCGTGCGGTGGGCGG 13497

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||   || |  || ||   ||||||  |  |||   | |  |  || | || ||| | |
Db      13496 GCGGTGGGCGTCGTGTTGGCGGGGGGAGTGGGCCAGGGGTGGGGGGGCGGGGCGTAGGCG 13437

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              ||  |     |||| |  ||| |  |  ||  |||||||  |  | || |||| |
Db      13436 GGTGCGGGGGGCGCGTAGAGGGCGAGGGCGGAGTGGGGGTGGGCCGGGGGCAGGG 13382
```

```
RESULT 12
SCO939128
LOCUS       SCO939128             296500 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 25/29.
ACCESSION   AL939128 AL023862 AL031013 AL031107 AL031350 AL031371 AL031515
            AL033505 AL034443 AL035161 AL079355 AL096825 AL645882
VERSION     AL939128.1  GI:24413907
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
   PUBMED   12000953
REFERENCE   2  (bases 1 to 296500)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
```

```
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 26, 2002 this sequence version replaced
            gi:20520750, gi:20520810, gi:20520753, gi:20520666, gi:20520757,
            gi:20520758, gi:20520668, gi:20520759, gi:20520814, gi:20520765,
            gi:20520767.
FEATURES             Location/Qualifiers
     source          1. .296500
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     gene            complement(144. .443)
                     /gene="SCO6480"
                     /note="synonym: SC9C7.16c"
     CDS             complement(144. .443)
                     /gene="SCO6480"
                     /note="SC9C7.16c, unknown, len: 99 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SC9C7.16c"
                     /protein_id="CAA22728.1"
                     /db_xref="GI:4154075"
                     /db_xref="GOA:Q9ZBJ4"
                     /db_xref="InterPro:IPR007138"
                     /db_xref="UniProt/TrEMBL:Q9ZBJ4"
                     /translation="MTYGCIASMKTRPGFREEVVSLLLSAADGLRDAGCELYVVGLSE
                     DDETTIWVTEVWRTREDHDASLELPAAKDAIGRAMPMLTGEFTKQEMSVAGGLGL"
     gene            complement(517. .1563)
                     /gene="SCO6481"
                     /note="synonym: SC9C7.17c"
     CDS             complement(517. .1563)
                     /gene="SCO6481"
                     /note="SC9C7.17c, conserved hypothetical protein, len:
                     348aa; similar to several hypothetical proteins eg.
                     SW:Y0BN_MYCTU from Mycobacterium tuberculosis (340 aa)
                     fasta scores; opt: 998, z-score: 1321.8, E(): 0, (47.6%
                     identity in 355 aa overlap)."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAA22729.1"
                     /db_xref="GI:4154076"
                     /db_xref="GOA:Q9ZBJ3"
                     /db_xref="InterPro:IPR002787"
                     /db_xref="InterPro:IPR006638"
                     /db_xref="InterPro:IPR007197"
                     /db_xref="UniProt/TrEMBL:Q9ZBJ3"
                     /translation="MRWENLTDSDHARAADTALFGADAVVTRTFDTPEFRGITFHEVR
                     ARSILNRVPGASRMPFEWTVNPYRGCTHACVYCFARKTHSYLDLDTGLGFDSQIVVKV
                     NAPEVLRRQLASRRWQGEHVAMGTNVDCYQRAEGRYRLMPGIIEALTERANPFSILTK
                     GTLILRDLDLLTRAARVTEVGISVSVGFTDAELWRTVEPGTPAPERRLEVVRALGERG
                     IGCGVLMAPVIPFLGDEPAQLRATVRAIAAAGATSVTPLVLHLRPGAREWFMAWLGRH
                     HPHLVRRYERLYADGAYAPKWYQRRITRQVHDLAREYGIGPARAGMPRRVAEPEHSRP
                     AEPAPPEPVQLSLI"
     gene            1800. .2243
                     /gene="SCO6482"
                     /note="synonym: SC9C7.18"
     CDS             1800. .2243
                     /gene="SCO6482"
                     /note="SC9C7.18, conserved hypothetical protein, len: 348
                     aa; similar to two hypothetical proteins e.g. TR:O69685
                     (EMBL:AL022121) Mycobacterium tuberculosis RV03718c or
                     MTV025.066c, 147 aa; fasta scores; opt: 475, z-score:
                     768.3, E(): 0, (52.1% identity in 146 aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAA22730.1"
                     /db_xref="GI:4154077"
                     /db_xref="UniProt/TrEMBL:Q9ZBJ2"
                     /translation="MAQVEATTERIVAADAETVFDTLADYSGTREKLLPEHFSEYEVR
                     EGGDGEGTLVHWKLQATSKRVRDCLLEVTEPTDGELVEKDRNSSMVTTWRVTPAGEGK
                     SRVVVLTTWQGAGGVGGFFEKTFAPKGLARIYDALLAKLAAEVEK"
     gene            2394. .4841
                     /gene="SCO6483"
                     /note="synonym: SC9C7.19"
     CDS             2394. .4841
                     /gene="SCO6483"
                     /note="SC9C7.19, probable efflux protein, len: 815aa;
                     similar to many, several of which are involved in toxin
                     /antibiotic efflux eg. TR:O69070 (EMBL:AF061335)
                     oxytetracycline exporter from Streptomyces rimosus (536
                     aa) fasta scores; opt: 1020, z-score: 879.9, E(): 0,
                     (36.9% identity in 542 aa overlap)."
                     /codon_start=1
                     /transl_table=11
                     /product="putative efflux protein"
                     /protein_id="CAA22731.1"
                     /db_xref="GI:4154078"
                     /db_xref="GOA:Q9ZBJ1"
                     /db_xref="InterPro:IPR000627"
                     /db_xref="InterPro:IPR001411"
                     /db_xref="InterPro:IPR001958"
                     /db_xref="InterPro:IPR004638"
                     /db_xref="InterPro:IPR007114"
                     /db_xref="InterPro:IPR008454"
                     /db_xref="InterPro:IPR011701"
                     /db_xref="UniProt/TrEMBL:Q9ZBJ1"
                     /translation="MGGITLVQDPVRDPVQDEPVVAAQAPPGPPVPGPDGERDAGLSP
                     RRVRLVFVGLMLALLLAALEQMIVATALPKIVGELHGLDKMSWAITAYLLTSTVGLPI
                     YGKLGDLFGRKGVFQFAILVFVVGSALAGRAQTMDQLIAFRAVQGVGAGGLMIGVQAI
                     IADIVPPRERARYMGLIGASFGLASVAGPLLGGYFTDHLSWRWCFYINVPFGLVTMAV
                     VAVVLKLPRPRVKPRLDILGALLLAAASTCLVLLTSWGGTEYAWGSRVILGLAAGAVA
                     ATLLFLVAEHFAPEPLIPLRLFRDSIFNVTALVGLVVGVALFGAASYLPTFLQMVDGA
                     SATESGLLMLPMMGGVVGASIISGQLISRTGHYRIHPILGSALSVVGMWLLSRLDADT
                     SRLHYSIWMAVLGAGIGLVMPVLVLAVQNSVRPTDLGTATSANNYFRQIGGSVGAAVF
                     GTLFAGRLTDALADRIPAEAGVGLPDAEAITPQLVHSLPPALRDAYIGAYADAMPRIF
                     LYLVPVLVLGLLIALFLKEKPLVSHNAPVTDPETAPHAPQPQPAAQVPPARSPYAGGI
                     PVCGTVQHPDGTVVPRAALTLIDVGGQQIGRGASGDDGRYALATPGSGAYVLIAAAGG
                     HQPQAVSVTVGERPVELDVVLGGAGRLAGSVMTADGSPVRDATVTLTNVHGEVVATTR
                     SGREGGYVITELVAGEYTLAASAPAFRPAALPVSVQAARETRQDVELAGGAVLRGTVR
                     ASGGRAVEDARVTLLDAAGNVVDTLTTGPDGTFRFVDLSSGEYTVIAAGYPPVATVLQ
                     VAGGGRTERDLQLGHED"
     gene            4911. .7493
                     /gene="SCO6484"
                     /note="synonym: SC9C7.20"
     CDS             4911. .7493
                     /gene="SCO6484"
                     /note="SC9C7.20, conserved hypothetical protein, len:
                     860aa; similar to many hypothetical proteins egs.
                     TR:O86875 (EMBL:U22894) from Streptomyces lividans (547
                     aa) fasta scores; opt: 1000, z-score: 738.2, E(): 0,
                     (40.0% identity in 497 aa overlap) and TR:O86807
                     (EMBL:AL031031) from Streptomyces coelicolor (916 aa)
                     fasta scores; opt: 358, z-score: 316.9, E(): 2.3e-10,
                     (27.2% identity in 779 aa overlap). Also contains a short
                     region of similarity to TR:Q53897 (EMBL:X60316) AbaA
                     regulatory locus from Streptomyces coelicolor (192 aa)
                     fasta scores; opt: 108, z-score: 223.2, E(): 3.8e-05,
                     (38.5% identity in 135 aa overlap)."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAA22732.1"
                     /db_xref="GI:4154079"
                     /db_xref="GOA:Q9ZBJ0"
                     /db_xref="InterPro:IPR000014"
                     /db_xref="InterPro:IPR001932"
                     /db_xref="InterPro:IPR003018"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="InterPro:IPR010822"
                     /db_xref="UniProt/TrEMBL:Q9ZBJ0"
                     /translation="MVTGHPRTVVTATQILRSRGQRGPDAMDHGTERDAPPGHRAGPD
                     TVPADPVTGRVPLAVVVVDRDGLVSHWSRGARRLFALSKEEAIGRPAPDLLPVSGVLP
                     EDAEPDGAGHDDSYRSYTAHDGPGPGLESSLDGRLSYPAAGRARLTVRADDRVDVLWW
                     AYPLVGPGPERLLVLAADADGLRRDYADGTDDSAGPAVVERIAPAFALHTDFPGADEL
```

```
                     ARRLPEILPSMSVGESARIVAQVLELGYPVLEFSQHDRVPVTPDWGVPRRAERRARRE
                     RAARALAAGEPVPDELRDEAEDLEYAAVRERLEFLNEVSGAIGTSLDLSRTIVEVSRA
                     VVPRFTDVAGTYLREQVVAGEGFPDGVPDTTTMWHRVALEHTDEPGRWDDVVPVGEAM
                     PFPAHTPFFQCMTTGEPVLVPRITEEMGHAIASQFEKRDIRPLITGRSMLLVPLKART
                     VVLGFMILLRHPERPVFNDMDRVTGAELAARAGLVLDNARMYTHQENVAETLQDSMLP
                     HIPARLAGCDIATRYLPGTLLGRVGGDWFDAVKLPGARTALVVGDVMGHGLNSAAMMG
                     QLRTAVQTMAGLDLPPAQLLRNLDDLAQRLGDTHLATCLYAVYDPIAGELHLANAGHI
                     PPVLVRAEDGASELIDLPTGAPIGVGGVPFESVRVTVAPGDRLVMCTDGLVEVRGEDI
                     GVGLATLCESAAHPAASMDDACDTIIRALGAGGGRKDDVALLMARLTGIEPDAVAEWR
                     LARDPAEVGRARAAVREQLYDWGLPKLVAPAELMVSELVTNAVRHSHARPVALRLVRA
                     DTLLCEVDDDDHELPALKSAGPGDETGRGLRVVSTLAREWGASRTGAGKTVWFELTLP
                     ARHR"
     gene            7711. .8433
                     /gene="SCO6485"
                     /note="synonym: SC9C7.21"
     CDS             7711. .8433
                     /gene="SCO6485"
                     /note="SC9C7.21, unknown, len: 240 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SC9C7.21"
                     /protein_id="CAA22733.1"
                     /db_xref="GI:4154080"
                     /db_xref="GOA:Q9ZBI9"
                     /db_xref="InterPro:IPR000051"

  Query Match             22.8%;  Score 67.2;  DB 1;  Length 296500;
  Best Local Similarity   53.9%;  Pred. No. 1.4e+02;
  Matches  138;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||| ||| | || | |  || ||||  |||  | | ||   |||  ||  |||  |||
Db     138651 GCGTCGGCCGCCTTCTCCGCGTGGTCGGCCGCGCTGTCGGCGGCGGCGGACGCGGACCGG 138710

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                |  ||||    | |||||   |   |   ||  ||  ||||||||||   | ||
Db     138711 GCGGCGGCGGCCGACCTGCGGGACAGGGTGGAGGCCCGGCTCGCGGCGGCCGTGGCCCGG 138770

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | |    ||    |  ||||    ||| |||||     ||||  ||| || |  ||
Db     138771 CCCGCGGCGGCCTCCGCGATGGCGGCCTGGCGCTGCGCCTCGGCGGCCTGGGACTGCGCG 138830

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||  ||||   |||| |  ||  |||||  |   |||||||| ||| |  | || |   |
Db     138831 GCACCGGAGGCCGCGGCGGCGTCGGCGGAAGCGCGCGCGGCGGCGGCCGAGTCGACCGCG 138890

Qy        241 GGGCCCACCTGCGCTT 256
              |  ||| |    || |
Db     138891 GCACCCGCGGCGGCGT 138906


RESULT 13
AC166628/c
LOCUS       AC166628               143360 bp    DNA     linear   HTG 03-AUG-2005
DEFINITION  Sorex araneus clone SA_Ba-587H22, WORKING DRAFT SEQUENCE, 6
            unordered pieces.
ACCESSION   AC166628
VERSION     AC166628.1  GI:71725472
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Sorex araneus (European shrew)
  ORGANISM  Sorex araneus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Insectivora; Soricidae;
            Soricinae; Sorex.
REFERENCE   1  (bases 1 to 143360)
  AUTHORS   Antonellis,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,
            Coleman,H., Engle,J., Franks,S., Fuksenko,T., Gestole,M.,
            Greene,A., Guan,X., Gupta,J., Gurson,N., Haghighi,P., Han,J.,
            Hansen,N., Ho,S.-L., Hu,P., Hunter,G., Hurle,B., Idol,J.R.,
            Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R.,
            Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Mojidi,H.A., Mullikin,J.C., Park,M.,
            Portnoy,M.E., Prasad,A., Puri,O., Rantz,K., Reddix-Dugue,N.,
            Sante,A., Schandler,K., Schueler,M.G., Sison,C., Stantripop,S.,
            Taye,A., Thomas,J.W., Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L.,
            Wetherby,K.D., Withers,T.R., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143360)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2005) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
COMMENT     -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                Center project name: lux
                Center clone name: 587H22
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 141076 bases at least Q40
                Consensus quality: 141317 bases at least Q30
                Consensus quality: 141436 bases at least Q20
                Insert size: 145000; agarose-fp
                Insert size: 142860; sum-of-contigs
                Quality coverage: 10.66x in Q20 bases; agarose-fp
                Quality coverage: 10.82x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3099: contig of 3099 bp in length
            *     3100     3199: gap of unknown length
            *     3200    11482: contig of 8283 bp in length
            *    11483    11582: gap of unknown length
            *    11583    28344: contig of 16762 bp in length
            *    28345    28444: gap of unknown length
            *    28445    52419: contig of 23975 bp in length
            *    52420    52519: gap of unknown length
            *    52520    89951: contig of 37432 bp in length
            *    89952    90051: gap of unknown length
            *    90052   143360: contig of 53309 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143360
                     /organism="Sorex araneus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:42254"
                     /clone="SA_Ba-587H22"
                     /clone_lib="SA_Ba"
     misc_feature    1. .3099
                     /note="assembly_fragment"
     gap             3100. .3199
                     /estimated_length=unknown
     misc_feature    3200. .11482
                     /note="assembly_fragment"
     gap             11483. .11582
                     /estimated_length=unknown
     misc_feature    11583. .28344
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             28345. .28444
                     /estimated_length=unknown
     misc_feature    28445. .52419
                     /note="assembly_fragment"
```

```
     gap             52420. .52519
                     /estimated_length=unknown
     misc_feature    52520. .89951
                     /note="assembly_fragment"
     gap             89952. .90051
                     /estimated_length=unknown
     misc_feature    90052. .143360
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
ORIGIN

  Query Match             22.7%;  Score 67;  DB 14;  Length 143360;
  Best Local Similarity   50.0%;  Pred. No. 1.9e+02;
  Matches  139;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              ||  |||||    |  || |  || |  | | | | |    ||  |  | |||   | |
Db      28239 CGCNGGCGGGGCGNCCGGCCCGGGCGGCGGGCGGCCGCGGCCGGNCGCCGGNCGGGCCGG 28180

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
                | ||||| | ||| | |   | |  || ||||| ||    |  |||| |  | |  | |
Db      28179 CGGGCCCGGGGCCGCGGGGCCGGNGGCGGGGCGCGGGGGCNGGNGCGGGCGCCCGGNCGC 28120

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
                  || ||| |  |  |    |  ||  |  ||| | || ||||    ||| || |||
Db      28119 CNGCGGCGGGGGGGGGGGGCGNGGGGCGCGGNGCGCNCNCGGGGGCNCNNGCGGGGCGCG 28060

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |||   | |||      || ||||||    || ||| ||  || |    || |   ||
Db      28059 GGGCGGGGGGGGCGGGGGGGCGGCGGGGCCGGGGGCGCCGGGGGGCGNCGCGCCCGGGG 28000

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGG 279
              |  |   | || |   | ||  ||    |  | |||||
Db      27999 GCGCGGGCCGCCCGGNCGGGGGCGCCGGGGNGGGGGG 27962


RESULT 14
PM7G11B/c
LOCUS       PM7G11B                 1094 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm7g11.b, sequence tagged site.
ACCESSION   AL685196
VERSION     AL685196.1  GI:19337723
KEYWORDS    STS.
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 1094)
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .1094
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm7g11.b"
ORIGIN

  Query Match             22.6%;  Score 66.8;  DB 10;  Length 1094;
  Best Local Similarity   51.2%;  Pred. No. 1.2e+03;
  Matches  149;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy          5 GGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCG 64
              || | | | ||| | || || |  | |   ||  | | ||||||||||  |  |   ||

Db        452 GGCCCGCCCGCGGGCCGCGGGCGGCGGGGGGGGGGGGGGCGCGGGCGGCCGGGGGTCGCG 393

Qy         65 TCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCC 124
                || |||  | |||  ||| |    || | |  | | | |||||   | ||
Db        392 CGCCCGCGCGGGCGCCCGCGGGGGCCCGGGGGGCGGGGGGGGCGGGGGGGGGNNGGGGGG 333

Qy        125 GCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCT 184
              |    | ||| |    | | | || |||| ||  |  |||||| |  || || ||||
Db        332 GGGGGGGGGGGGGCGCCGGGGGCCCGCGCGCGGCCCGGCGGGCGCGGCCGGGGCGCGCNG 273

Qy        185 CGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGC 244
              | | |  |||     | |  | ||| |||    || |  ||   ||  | ||   | |
Db        272 CCGCCCCGGCGCGGCGGCGCCCCGGCCGCGCGCGGGGGGGGGCGGGGGGGGCGGGCGCGG 213

Qy        245 CCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |  | | ||    | || ||    |   || || ||  | ||  | || | |
Db        212 CGCCGGGGGCCCGCGGGGCCCCCCGCNCGGGGNGGGGGGGCGCGGGCGGGG 162


RESULT 15
AC084804/c
LOCUS       AC084804               219952 bp    DNA     linear   HTG 15-MAY-2002
DEFINITION  Mus musculus strain C57BL6/J clone RP23-314K21, WORKING DRAFT
            SEQUENCE, 46 unordered pieces.
ACCESSION   AC084804
VERSION     AC084804.10  GI:17488575
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 219952)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 219952)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2000) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
COMMENT     On Dec 11, 2001 this sequence version replaced gi:13560369.
            ----------------Genome Center
            Center:  Harvard Partners Genome Center
            Center Code:  HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact: hpgc@mendel.mgh.harvard.edu
            ----------------Summary Statistics
            Center project name:  AEZ
            Sequencing vector:  pUC18; L08752
            Chemistry:  Dye-terminator Big Dye; 100%
            *Consensus quality:  207914 at least Q20
            *Consensus quality:  205077 at least Q30
            *Consensus quality:  200559 at least Q40
            Estimated insert size:   agarose-FP - N/A
            **Estimated insert size: 219052 - sum-of-contigs
            Quality coverage:  agarose-FP - N/A
            Quality coverage:  10.7 x in Q20 bases; sum-of-contigs estimation
            ----------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 46 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    26658: contig of 26658 bp in length
```

```
*    26659    26678: gap of unknown length
*    26679    53321: contig of 26643 bp in length
*    53322    53341: gap of unknown length
*    53342    75821: contig of 22480 bp in length
*    75822    75841: gap of unknown length
*    75842    99596: contig of 23755 bp in length
*    99597    99616: gap of unknown length
*    99617   120496: contig of 20880 bp in length
*   120497   120516: gap of unknown length
*   120517   132931: contig of 12415 bp in length
*   132932   132951: gap of unknown length
*   132952   144027: contig of 11076 bp in length
*   144028   144047: gap of unknown length
*   144048   153071: contig of 9024 bp in length
*   153072   153091: gap of unknown length
*   153092   162146: contig of 9055 bp in length
*   162147   162166: gap of unknown length
*   162167   167980: contig of 5814 bp in length
*   167981   168000: gap of unknown length
*   168001   173254: contig of 5254 bp in length
*   173255   173274: gap of unknown length
*   173275   179246: contig of 5972 bp in length
*   179247   179266: gap of unknown length
*   179267   180707: contig of 1441 bp in length
*   180708   180727: gap of unknown length
*   180728   184387: contig of 3660 bp in length
*   184388   184407: gap of unknown length
*   184408   189237: contig of 4830 bp in length
*   189238   189257: gap of unknown length
*   189258   192994: contig of 3737 bp in length
*   192995   193014: gap of unknown length
*   193015   197265: contig of 4251 bp in length
*   197266   197285: gap of unknown length
*   197286   197589: contig of 304 bp in length
*   197590   197609: gap of unknown length
*   197610   199581: contig of 1972 bp in length
*   199582   199601: gap of unknown length
*   199602   201006: contig of 1405 bp in length
*   201007   201026: gap of unknown length
*   201027   202394: contig of 1368 bp in length
*   202395   202414: gap of unknown length
*   202415   203415: contig of 1001 bp in length
*   203416   203435: gap of unknown length
*   203436   204824: contig of 1389 bp in length
*   204825   204844: gap of unknown length
*   204845   205380: contig of 536 bp in length
*   205381   205400: gap of unknown length
*   205401   207598: contig of 2198 bp in length
*   207599   207618: gap of unknown length
*   207619   207875: contig of 257 bp in length
*   207876   207895: gap of unknown length
*   207896   208001: contig of 106 bp in length
*   208002   208021: gap of unknown length
*   208022   208140: contig of 119 bp in length
*   208141   208160: gap of unknown length
*   208161   208450: contig of 290 bp in length
*   208451   208470: gap of unknown length
*   208471   209852: contig of 1382 bp in length
*   209853   209872: gap of unknown length
*   209873   210569: contig of 697 bp in length
*   210570   210589: gap of unknown length
*   210590   210757: contig of 168 bp in length
*   210758   210777: gap of unknown length
*   210778   211360: contig of 583 bp in length
*   211361   211380: gap of unknown length
*   211381   211510: contig of 130 bp in length
*   211511   211530: gap of unknown length
*   211531   212447: contig of 917 bp in length
*   212448   212467: gap of unknown length
*   212468   213049: contig of 582 bp in length
*   213050   213069: gap of unknown length
*   213070   213333: contig of 264 bp in length
*   213334   213353: gap of unknown length
*   213354   214164: contig of 811 bp in length
*   214165   214184: gap of unknown length
*   214185   214400: contig of 216 bp in length
*   214401   214420: gap of unknown length
*   214421   214957: contig of 537 bp in length
*   214958   214977: gap of unknown length
*   214978   215140: contig of 163 bp in length
*   215141   215160: gap of unknown length
*   215161   215345: contig of 185 bp in length
*   215346   215365: gap of unknown length
*   215366   216233: contig of 868 bp in length
*   216234   216253: gap of unknown length
*   216254   216819: contig of 566 bp in length
*   216820   216839: gap of unknown length
*   216840   219659: contig of 2820 bp in length
*   219660   219679: gap of unknown length
*   219680   219952: contig of 273 bp in length.
FEATURES             Location/Qualifiers
     source          1. .219952
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /clone="RP23-314K21"
                     /sex="male"
     misc_feature    1. .26658
                     /note="assembly_name:Contig192"
     gap             26659. .26678
                     /estimated_length=unknown
     misc_feature    26679. .53321
                     /note="assembly_name:Contig191"
     gap             53322. .53341
                     /estimated_length=unknown
     misc_feature    53342. .75821
                     /note="assembly_name:Contig190"
     gap             75822. .75841
                     /estimated_length=unknown
     misc_feature    75842. .99596
                     /note="assembly_name:Contig189"
     gap             99597. .99616
                     /estimated_length=unknown
     misc_feature    99617. .120496
                     /note="assembly_name:Contig188"
     gap             120497. .120516
                     /estimated_length=unknown
     misc_feature    120517. .132931
                     /note="assembly_name:Contig187"
     gap             132932. .132951
                     /estimated_length=unknown
     misc_feature    132952. .144027
                     /note="assembly_name:Contig186"
     gap             144028. .144047
                     /estimated_length=unknown
     misc_feature    144048. .153071
                     /note="assembly_name:Contig185"
     gap             153072. .153091
                     /estimated_length=unknown
     misc_feature    153092. .162146
                     /note="assembly_name:Contig184"
     gap             162147. .162166
                     /estimated_length=unknown
     misc_feature    162167. .167980
                     /note="assembly_name:Contig183"
     gap             167981. .168000
                     /estimated_length=unknown
     misc_feature    168001. .173254
                     /note="assembly_name:Contig182"
     gap             173255. .173274
                     /estimated_length=unknown
     misc_feature    173275. .179246
                     /note="assembly_name:Contig181"
     gap             179247. .179266
                     /estimated_length=unknown
```

misc_feature 179267. .180707
/note="assembly_name:Contig180"

Query Match 22.5%; Score 66.4; DB 14; Length 219952;
Best Local Similarity 51.0%; Pred. No. 1.8e+02;
Matches 148; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```
Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              | | ||  || | | | | || || ||   |   | | | | | ||| |  ||| | ||
Db     188869 GGGCGGCGGGGCGGGGGGGCGGGGGCCGGGGCGNGCCGGGGGGGGCGCGGCGCGGGGCGG 188810

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                 |    || | ||  ||||    |  || | | | || |||||||||| ||||||
Db     188809 CGGGGGGGGGGGGCGGGGCGGCGCCGGCGGGGGGGGCGGGGCGCGGCGGGCCGCGGGCNG 188750

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | || |  | ||| |     |   |  ||| |   | |  ||  || | |||   || |
Db     188749 CCCCNCNNGCGGGGGGGCGGCGCGGGGCGGGGGCGGGGGGCGGCGGGGGGCGGCGGGCGG 188690

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||  ||     | | |     | || |||||| ||| || | | ||| | |||   |   |
Db     188689 GCGGCGCGGGGGGGGGGGCGGCGGCCGGGCGCCGGGGCCGGGGCGGCCGGGCGAGCAGCG 188630

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTG 290
              ||  |  ||  | |  ||  | ||   |   |  ||  |  |   ||||
Db     188629 GGCGCGCCCCCCCCCCCCCCCGCCCCCCCGCCCNCGGTCGTGGTAGGGGTG 188580
```

Search completed: April 28, 2006, 02:39:46
Job time : 1636.62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 06:17:02 ; Search time 142.252 Seconds
(without alignments)
13821.091 Million cell updates/sec

Title: US-10-729-802-1_COPY_77000_77294
Perfect score: 295
Sequence: 1 gcgagggcggtctgcgcgtc.....ggggccgagcaggtgcagcg 295

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		295	100.0	76994	12	ADP64454	Adp64454 Sorangium
2	c	70.4	23.9	73882	13	ADS73531	Ads73531 tcp gene
3	c	69	23.4	1149	12	ADO00480	Ado00480 Novel hum
4	c	69	23.4	1149	12	ADN98911	Adn98911 Novel hum
5	c	69	23.4	1470	12	ADP28824	Adp28824 Human sec
6		66	22.4	138203	14	ADZ51725	Adz51725 FR-008 po
7	c	64.4	21.8	1338	14	ACL67140	Acl67140 M. xanthu
8	c	64.4	21.8	20113	14	ACL64710	Acl64710 M. xanthu
9		63.8	21.6	1047	8	AAL61180	Aal61180 Actinosyn
10		63.8	21.6	82746	8	AAL61224	Aal61224 Actinosyn
11		63.6	21.6	3959	6	ABI99206	Abi99206 Mouse isc
12	c	63.4	21.5	103599	4	ABX04971	Abx04971 S. cinnam
13		62.4	21.2	71989	3	AAA29349	Aaa29349 Sorangium
14		62	21.0	3403	14	ADX26149	Adx26149 Novel cel
15	c	61.8	20.9	2184	8	ACA26420	Aca26420 Prokaryot
16	c	61.4	20.8	114955	2	AAX53491	Aax53491 Human ade
17	c	61.2	20.7	58857	3	AAA58471	Aaa58471 Nucleotid
18	c	61.2	20.7	109519	5	AAS08693	Aas08693 Micromono
19	c	60.8	20.6	4257	2	AAV68520	Aav68520 The nucle
20	c	60.8	20.6	4257	2	AAV10362	Aav10362 Infected
21		60.8	20.6	12001	2	AAQ76213	Aaq76213 HSV L/ST
22		60.8	20.6	37185	13	ADW39406	Adw39406 Herpes si
23	c	60.8	20.6	40772	13	ADW39403	Adw39403 Herpes si
24	c	60.4	20.5	9975	8	AAL61173	Aal61173 Actinosyn
25		60.2	20.4	37856	3	AAA11992	Aaa11992 S. cellul
26	c	60	20.3	4667	12	ADQ22939	Adq22939 Human sof
27		60	20.3	114955	2	AAX53491	Aax53491 Human ade
28		59.8	20.3	1162	2	AAT69796	Aat69796 Human neu
29		59.8	20.3	1162	2	AAX75760	Aax75760 Human neu
30		59.8	20.3	1162	6	ABL69705	Abl69705 Prostate
31		59.8	20.3	2325	12	ADQ23872	Adq23872 Human sof
32		59.8	20.3	3750	13	ADQ86035	Adq86035 Human tum
33		59.8	20.3	3774	10	ADK11436	Adk11436 Human neu
34		59.8	20.3	3774	12	ADQ19719	Adq19719 Human sof
35		59.8	20.3	3774	13	ADW14258	Adw14258 Human neu
36		59.8	20.3	3774	14	ADX25912	Adx25912 Novel cel
37		59.8	20.3	4037	4	AAK52942	Aak52942 Human pol
38		59.8	20.3	7713	4	AAK51958	Aak51958 Human pol
39	c	59.6	20.2	1239	8	ACA25997	Aca25997 Prokaryot
40	c	59.6	20.2	3957	5	AAA09686	Aaa09686 HSV-2 imm
41	c	59.6	20.2	3957	10	ADG75121	Adg75121 Human her
42		59.6	20.2	154746	6	AAD25519	Aad25519 Human her
43	c	59.6	20.2	154746	6	AAD25519	Aad25519 Human her
44	c	59.2	20.1	1677	8	ADA70177	Ada70177 Rice gene
45	c	59.2	20.1	2302	12	ADP03322	Adp03322 Rice grai

ALIGNMENTS

RESULT 1
ADP64454
ID ADP64454 standard; DNA; 76994 BP.
XX
AC ADP64454;
XX
DT 09-SEP-2004 (first entry)
XX
DE Sorangium cellulosum disorazole polyketide synthase gene cluster DNA.
XX
KW disorazole polyketide synthase; disorazole PKS; polyketide synthase; PKS;
KW enzyme; dszA; dszB; dszC; dszD; disorazole; antimicrobial;
KW tubulin polymerisation inhibitor; microtubule decay;
KW disorazole polyketide; disorazole polyketide synthase gene cluster; gene;
KW ds.
XX
OS Polyangium cellulosum.
XX
PN WO2004053065-A2.
XX
PD 24-JUN-2004.
XX
PF 05-DEC-2003; 2003WO-US038500.
XX
PR 06-DEC-2002; 2002US-0431272P.
PR 17-MAR-2003; 2003US-0455521P.
PR 23-APR-2003; 2003US-0465038P.
PR 22-MAY-2003; 2003US-0473311P.
PR 02-JUL-2003; 2003US-0484934P.
PR 20-OCT-2003; 2003US-0512892P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Julien B, Reid RC;
XX
DR WPI; 2004-468841/44.
XX
PT Novel isolated recombinant polynucleotide comprising nucleotide sequence
PT encoding disorazole polyketide synthase protein or fragment comprising
PT domain of polyketide synthase protein, useful in human and veterinary
PT medicine.
XX

```
PS   Claim 2; SEQ ID NO 1; 72pp; English.
XX
CC   The present invention describes an isolated recombinant polynucleotide
CC   (I) comprising a nucleotide sequence encoding a disorazole polyketide
CC   synthase (PKS) protein or a fragment comprising one or more domains of
CC   the PKS. Also described: (1) vector (II) comprising (I); (2) recombinant
CC   host cell (III) comprising (II), or (I) integrated into the cell
CC   chromosomal DNA; (3) chimeric PKS (IV) comprising one or more domain of a
CC   disorazole PKS; (4) modified functional disorazole PKS (V) differing from
CC   the native disorazole PKS by the inactivation of one or more domain of
CC   the disorazole PKS and/or addition of one or more domain of a non-
CC   disorazole PKS; (5) cell comprising (IV) or (V); (6) recombinant
CC   expression system (VI) capable of producing a disorazole synthase domain
CC   in a host cell, where the system comprises an encoding sequence for a
CC   disorazole polyketide synthase domain, and the encoding sequence is
CC   operably linked to control sequences effective in the cell to produce RNA
CC   that is translated into the domain; (7) host cell modified to contain
CC   (VI); (8) recombinant Sorangium cellulosum (Polyangium cellulosum) cell
CC   in which a dszA, dszB, dszC or dszD gene is disrupted to reduce or
CC   eliminate production of disorazole; (9) an isolated polypeptide encoded
CC   by (I); (10) hybrid polyketide synthase comprising one or more
CC   polypeptides of a disorazole PKS and one or more polypeptides of a non-
CC   disorazole PKS; and (11) recombinant DNA molecule, comprising a sequence
CC   of at least 200 base pairs with a sequence identical or substantially
CC   identical to a protein encoding region of the 77294 base pair sequence of
CC   SEQ ID NO:1 (S1). (I) has antimicrobial activities, and can be used as an
CC   inhibitor of tubulin polymerisation, and induces decay of microtubules.
CC   (I) is useful for producing disorazole polyketides. (III) is useful for
CC   producing a polyketide, which involves growing (III) under conditions,
CC   where a polyketide synthesised by a PKS comprising a protein encoded by
CC   the polynucleotide molecule is produced in the cell. (I) is useful for
CC   producing libraries of PKSs, where the polyketide producing colonies are
CC   identified and isolated, and the produced polyketides are used
CC   collectively in a panel to represent a library or assessed individually
CC   for activity. The disorazole polyketide produced by (I) is useful in
CC   molecular biology, chemistry, recombinant DNA technology, human and
CC   veterinary medicine and agricultural applications. The disorazole
CC   polyketide produced by (I) is formulated as a pharmaceutical composition,
CC   useful for its activity against microorganisms, in a mammal. The present
CC   sequence represents the Sorangium cellulosum (Polyangium cellulosum)
CC   disorazole polyketide synthase gene cluster nucleotide sequence, which is
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 76994 BP; 10739 A; 28790 C; 26877 G; 10587 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 295;  DB 12;  Length 76994;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-34;
  Matches  295;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76700 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 76759

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76760 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 76819

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76820 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 76879

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76880 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 76939

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76940 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 76994


RESULT 2
ADS73531/c
ID   ADS73531 standard; cDNA; 73882 BP.
XX
AC   ADS73531;
XX
DT   02-DEC-2004  (first entry)
XX
DE   tcp gene cluster.
XX
KW   ss; gene; tcp ; gene cluster ; biosynthesis ; glycopeptide ; teicoplanin;
KW   antibiotic; bal; cep; com ; sta ; 4-hydroxy-phenylglycine ; 3;
KW   5-dihydroxy-phenylglycine; heptapeptide skeleton ; cross-linking ;
KW   aromatic residue; N-acyl-beta-D-glucosamine ; core structure ; export;
KW   resistance; chlorination ; alpha-hydroxylation ; mannosyl attachment.
XX
OS   Actinoplanes teichomyceticus; ATCC31121.
XX
FH   Key             Location/Qualifiers
FT   CDS             652. .1995
FT                   /*tag=  a
FT                   /label= ORF1
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin"
FT                   /product= "MurF"
FT   CDS             2017. .3234
FT                   /*tag=  b
FT                   /label= ORF2
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin"
FT   CDS             3227. .4264
FT                   /*tag=  c
FT                   /label= ORF3
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin"
FT   CDS             4264. .4872
FT                   /*tag=  d
FT                   /label= ORF4
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin"
FT   CDS             complement(5634. .6173)
FT                   /*tag=  e
FT                   /label= ORF5
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin"
FT   CDS             complement(6286. .7377)
FT                   /*tag=  f
FT                   /label= ORF6
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin
FT                   and required for regulatingexpression of one or more of
FT                   the genes of tcp gene cluster"
FT   CDS             complement(7385. .8065)
FT                   /*tag=  g
FT                   /label= ORF7
FT                   /note= "Required for regulatingexpression of one or more
FT                   of the genes of tcp gene cluster"
FT   CDS             8924. .10105
FT                   /*tag=  h
FT                   /label= ORF8
FT                   /note= "Required for the addition and formation of N-acyl
FT                   -beta-glucosamine residues"
FT   CDS             10597. .16827
FT                   /*tag=  i
FT                   /label= ORF9
FT                   /note= "Required for synthesis of the heptapeptide
FT                   skeleton"
FT   CDS             16824. .20009
FT                   /*tag=  j
FT                   /label= ORF10
FT                   /note= "Required for synthesis of the heptapeptide
FT                   skeleton"
FT   CDS             20053. .32256
FT                   /*tag=  k
FT                   /label= ORF11
FT                   /note= "Required for synthesis of the heptapeptide
```

```
FT                   skeleton"
FT   CDS             32276. .37873
FT                   /*tag=  l
FT                   /label= ORF12
FT                   /note= "Required for synthesis of the heptapeptide
FT                   skeleton"
FT   CDS             37886. .38095
FT                   /*tag=  m
FT                   /label= ORF13
FT   CDS             38401. .39222
FT                   /*tag=  n
FT                   /label= ORF14
FT   CDS             39268. .41049
FT                   /*tag=  o
FT                   /label= ORF15
FT                   /note= "Required for attachment of mannosyl residue"
FT   CDS             41343. .43289
FT                   /*tag=  p
FT                   /label= ORF16
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin"
FT   CDS             43373. .43585
FT                   /*tag=  q
FT                   /label= ORF17
FT   CDS             43696. .44871
FT                   /*tag=  r
FT                   /label= ORF18
FT                   /note= "Required for cross-linking of aromatic residues
FT                   atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT                   teicoplanin"
FT   CDS             44894. .46048
FT                   /*tag=  s
FT                   /label= ORF19
FT                   /note= "Required for cross-linking of aromatic residues
FT                   atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT                   teicoplanin"
FT   CDS             46038. .47234
FT                   /*tag=  t
FT                   /label= ORF20
FT                   /note= "Required for cross-linking of aromatic residues
FT                   atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT                   teicoplanin"
FT   CDS             47484. .49004
FT                   /*tag=  u
FT                   /label= ORF21
FT                   /note= "Required for the chlorination of the aromatic
FT                   residues"
FT   CDS             49166. .50344
FT                   /*tag=  v
FT                   /label= ORF22
FT                   /note= "Required for cross-linking of aromatic residues
FT                   atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT                   teicoplanin"
FT   CDS             50461. .51687
FT                   /*tag=  w
FT                   /label= ORF23
FT                   /note= "Required for the addition and formation of N-acyl
FT                   -beta-glucosamine residues"
FT   CDS             51833. .52804
FT                   /*tag=  x
FT                   /label= ORF24
FT                   /note= "Required for the addition and formation of N-acyl
FT                   -beta-glucosamine residues"
FT   CDS             52948. .54540
FT                   /*tag=  y
FT                   /label= ORF25
FT                   /note= "Required for the beta-hydroxylation of the Tyr
FT                   residues"
FT   CDS             54757. .56553
FT                   /*tag=  z
FT                   /label= ORF26
FT                   /note= "Required for the addition and formation of N-acyl
FT                   -beta-glucosamine residues"
FT   CDS             56924. .57979
FT                   /*tag=  aa
FT                   /label= ORF27
FT   CDS             58746. .59735
FT                   /*tag=  ab
FT                   /label= ORF28
FT                   /note= "Required for regulatingexpression of one or more
FT                   of the genes of tcp gene cluster"
FT   CDS             60477. .62855
FT                   /*tag=  ac
FT                   /label= ORF29
FT                   /note= "Required for regulatingexpression of one or more
FT                   of the genes of tcp gene cluster"
FT   CDS             63036. .64157
FT                   /*tag=  ad
FT                   /label= ORF30
FT                   /note= "Required for synthesis of the 3,5-dihydroxy-
FT                   phenylglycine residues"
FT   CDS             64154. .64819
FT                   /*tag=  ae
FT                   /label= ORF31
FT                   /note= "Required for synthesis of the 3,5-dihydroxy-
FT                   phenylglycine residues"
FT   CDS             64816. .66117
FT                   /*tag=  af
FT                   /label= ORF32
FT                   /note= "Required for synthesis of the 3,5-dihydroxy-
FT                   phenylglycine residues"
FT   CDS             66146. .66952
FT                   /*tag=  ag
FT                   /label= ORF33
FT                   /note= "Required for synthesis of the 3,5-dihydroxy-
FT                   phenylglycine residues"
FT   CDS             66991. .68352
FT                   /*tag=  ah
FT                   /label= ORF34
FT                   /note= "Required for export of teicoplanin outside of the
FT                   cytoplasm, and for coferring resistance to teicoplanin"
FT   CDS             68349. .68945
FT                   /*tag=  ai
FT                   /label= ORF35
FT   CDS             69083. .70180
FT                   /*tag=  aj
FT                   /label= ORF36
FT                   /note= "Required for synthesis of the 4-hydroxy-
FT                   phenylglycine residues, and required for synthesis of the
FT                   3,5-dihydroxy-phenylglycine residues"
FT   CDS             70266. .71321
FT                   /*tag=  ak
FT                   /label= ORF37
FT                   /note= "Required for synthesis of the 4-hydroxy-
FT                   phenylglycine residues"
FT   CDS             71318. .72412
```

```
  Query Match             23.9%;  Score 70.4;  DB 13;  Length 73882;
  Best Local Similarity   54.7%;  Pred. No. 0.023;
  Matches  162;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    1;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              | |  | |||  ||     | || |||| ||      ||||||  |    ||| ||||
Db      35827 GTGCACCGTCACCGACAGCACGTGCTCCTTCTCCGACAGCGCGAAGAGCTCGCACCGCCA 35768

Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCG---CGGCCC 119
              || |  | |   |  | || ||  |||  || ||    || ||||   |||    ||
Db      35767 CGGCATGTCCCGGGTGAGGTCGAACAGCTGCTCGCCCCGCTCGGCCAGCCGTGCGGGGAG 35708

Qy        120 CCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAG 179
                |  ||||  |||   || | ||| | | || ||  || ||  | ||||   |||  |
Db      35707 GTCCTCCTCGGTGGCCGGGACCGGCGTCAGCCGGACCGGCGCGGCGTCGTGGATGTGCTG 35648

Qy        180 CGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGAT 239
              |||  | |||||||  | |||  ||  | | ||  |    ||  | || |  |||
Db      35647 GTGGACGGTCTGCGCGTCGCCCGGGAAGGTGGTCCGCAGGATCTCGTGCCGGGCCGCGAC 35588
```

```
Qy        240 GGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |  || || |  ||| |||||| | |   ||||  |  ||||| ||||| ||||||
Db      35587 GTCGCGCAGCGCCGCCTCCAGGGCGGCCACGTCCAGCCGGCCGCGCAGGCGCAGCG 35532
```

RESULT 3
ADO00480/c
ID ADO00480 standard; cDNA; 1149 BP.
XX
AC ADO00480;
XX
DT 29-JUL-2004 (first entry)
XX
DE Novel human cDNA sequence #1295.
XX
KW ds; gene; anti-inflammatory; dermatological; neuroprotective;
KW immunomodulator; antibacterial; virucide; antipsoriatic; cytostatic;
KW gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW psoriasis; diabetes; early aging; hormonal imbalance;
KW ischemic heart disease; ulcerative colitis.
XX
OS Homo sapiens.
XX
PN WO2004038003-A2.
XX
PD 06-MAY-2004.
XX
PF 24-OCT-2003; 2003WO-US033947.
XX
PR 25-OCT-2002; 2002US-0421061P.
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421552P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422177P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 15-NOV-2002; 2002US-0426394P.
PR 15-NOV-2002; 2002US-0426430P.
PR 15-NOV-2002; 2002US-0426916P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 27-NOV-2002; 2002US-0429302P.
PR 27-NOV-2002; 2002US-0429326P.
PR 27-NOV-2002; 2002US-0429651P.
PR 04-DEC-2002; 2002US-0430645P.
PR 04-DEC-2002; 2002US-0430651P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430668P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430937P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433251P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433316P.
PR 13-DEC-2002; 2002US-0433318P.
PR 23-DEC-2002; 2002US-0436236P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 17-JAN-2003; 2003US-0440821P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463708P.
PR 18-APR-2003; 2003US-0463716P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467199P.
PR 02-MAY-2003; 2003US-0467201P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 19-MAY-2003; 2003US-0471336P.
PR 22-MAY-2003; 2003US-0472420P.
PR 22-MAY-2003; 2003US-0472430P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476621P.
PR 09-JUN-2003; 2003US-0476632P.
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485217P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486446P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;
PI Wong JGP, Wu G, Zhang H, Zeng C;
XX
DR WPI; 2004-365511/34.
DR P-PSDB; ADN99695.
XX
PT New nucleic acid molecules, useful in preparing a composition for
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT ulcerative colitis.
XX
PS Claim 1; SEQ ID NO 2079; 532pp; English.
XX
CC The invention relates to a nucleic acid molecule comprising a
CC polynucleotide sequence or its complement that encodes a polypeptide. The
CC nucleic acid is useful in preparing a composition for treating or
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC heart disease or ulcerative colitis. This sequence corresponds to a
CC nucleic acid of the invention.
XX
SQ Sequence 1149 BP; 137 A; 492 C; 390 G; 130 T; 0 U; 0 Other;

 Query Match 23.4%; Score 69; DB 12; Length 1149;
 Best Local Similarity 55.7%; Pred. No. 0.11;
 Matches 132; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy         45 GCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGC 104
              || | | ||||  |||  || ||||| | ||  |  ||| |  ||||  || || ||||
Db        462 GCCGCCCGCGCCGCGCCGCGACCCGGGGCCGGGGACCGCGCGGGGGCGGCCGGGGGGCGC 403

Qy        105 GGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCG 164
              | || |||||||||     |    || |  ||  ||   |||  | | |  || |  || |
Db        402 CGGGGGCCGCGGCCGGGGCGCGCCGGGACCGGGGACGCGCGGGCCGGGGCCGGGTCCGGG 343

Qy        165 GGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCC 224
              | | ||||| ||    |   || | |   | |  |||||  |||  | ||  ||| |||
Db        342 CGAGGGCGCGCGGGCGCCGCGGGCCGGGCGGGCGCCGGGGCGGGGGCGGGGACCGGGGCC 283

Qy        225 GGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCC 281
              || || | |  |  | |  ||  |    | ||   |  |  |||  || || |||||||
Db        282 GGGCAAGGCTCCCGTCAGGGGCCGGCGGGCGGGCGCGCGCGGGGCGCCGGGGGGCC 226
```

RESULT 4
ADN98911/c
ID ADN98911 standard; cDNA; 1149 BP.
XX
AC ADN98911;

XX
DT 29-JUL-2004 (first entry)
XX
DE Novel human cDNA sequence #511.
XX
KW ds; gene; anti-inflammatory; dermatological; neuroprotective;
KW immunomodulator; antibacterial; virucide; antipsoriatic; cytostatic;
KW gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW psoriasis; diabetes; early aging; hormonal imbalance;
KW ischemic heart disease; ulcerative colitis.
XX
OS Homo sapiens.
XX
PN WO2004038003-A2.
XX
PD 06-MAY-2004.
XX
PF 24-OCT-2003; 2003WO-US033947.
XX
PR 25-OCT-2002; 2002US-0421061P.
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421552P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422177P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 15-NOV-2002; 2002US-0426394P.
PR 15-NOV-2002; 2002US-0426430P.
PR 15-NOV-2002; 2002US-0426916P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 27-NOV-2002; 2002US-0429302P.
PR 27-NOV-2002; 2002US-0429326P.
PR 27-NOV-2002; 2002US-0429651P.
PR 04-DEC-2002; 2002US-0430645P.
PR 04-DEC-2002; 2002US-0430651P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430668P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430937P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433251P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433316P.
PR 13-DEC-2002; 2002US-0433318P.
PR 23-DEC-2002; 2002US-0436236P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 17-JAN-2003; 2003US-0440821P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463708P.
PR 18-APR-2003; 2003US-0463716P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467199P.
PR 02-MAY-2003; 2003US-0467201P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 19-MAY-2003; 2003US-0471336P.
PR 22-MAY-2003; 2003US-0472420P.
PR 22-MAY-2003; 2003US-0472430P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476621P.
PR 09-JUN-2003; 2003US-0476632P.
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485217P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486446P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;
PI Wong JGP, Wu G, Zhang H, Zeng C;
XX
DR WPI; 2004-365511/34.
DR P-PSDB; ADN99695.
XX
PT New nucleic acid molecules, useful in preparing a composition for
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT ulcerative colitis.
XX
PS Claim 1; SEQ ID NO 511; 532pp; English.
XX
CC The invention relates to a nucleic acid molecule comprising a
CC polynucleotide sequence or its complement that encodes a polypeptide. The
CC nucleic acid is useful in preparing a composition for treating or
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC heart disease or ulcerative colitis. This sequence corresponds to a
CC nucleic acid of the invention.
XX
SQ Sequence 1149 BP; 137 A; 492 C; 390 G; 130 T; 0 U; 0 Other;

Query Match 23.4%; Score 69; DB 12; Length 1149;
Best Local Similarity 55.7%; Pred. No. 0.11;
Matches 132; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy         45 GCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGC 104
              || | | ||||  |||  || ||||| | ||  |   ||| |  ||||  || || ||||
Db        462 GCCGCCCGCGCCGCGCCGCGACCCGGGGCCGGGGACCGCGCGGGGGCGGCCGGGGGGCGC 403

Qy        105 GGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCG 164
               | || ||||||||     |    || | ||  ||  ||| | | |  || |  || |
Db        402 CGGGGGCCGCGGCCGGGGCGCGCCGGGACCGGGGACGCGCGGGCCGGGGCCGGGTCCGGG 343

Qy        165 GGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCC 224
               | | ||||| ||     |   || | |   | |  |||||  |||  | ||  ||| |||
Db        342 CGAGGGCGCGCGGGCGCCGCGGGCCGGGCGGGCGCCGGGGCGGGGGCGGGGACCGGGGCC 283

Qy        225 GGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCC 281
              || || | |  |  |  ||  |    | ||   |  |  |||  || || |||||||
Db        282 GGGCAAGGCTCCCGTCAGGGGCCGGCGGGCGGGCGCGCGCGGGGCGCCGGGGGGGCC 226
```

RESULT 5
ADP28824/c
ID ADP28824 standard; DNA; 1470 BP.
XX
AC ADP28824;
XX
DT 12-AUG-2004 (first entry)
XX
DE Human secreted protein encoding sequence SEQ ID #822.
XX
KW Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
KW cancer; inflammatory; immune; ds; human secreted protein.
XX
OS Homo sapiens.
XX
PN WO2004035732-A2.

```
XX
PD   29-APR-2004.
XX
PF   28-AUG-2003; 2003WO-US026780.
XX
PR   29-AUG-2002; 2002US-0406576P.
PR   29-AUG-2002; 2002US-0406579P.
PR   29-AUG-2002; 2002US-0406585P.
PR   29-AUG-2002; 2002US-0406588P.
PR   29-AUG-2002; 2002US-0406608P.
PR   29-AUG-2002; 2002US-0406611P.
PR   29-AUG-2002; 2002US-0406612P.
PR   29-AUG-2002; 2002US-0406616P.
PR   29-AUG-2002; 2002US-0406640P.
PR   29-AUG-2002; 2002US-0406642P.
PR   29-AUG-2002; 2002US-0406646P.
PR   29-AUG-2002; 2002US-0406653P.
PR   29-AUG-2002; 2002US-0406655P.
PR   29-AUG-2002; 2002US-0406666P.
PR   17-SEP-2002; 2002US-0410946P.
PR   17-SEP-2002; 2002US-0410947P.
PR   17-SEP-2002; 2002US-0410948P.
PR   17-SEP-2002; 2002US-0410949P.
PR   17-SEP-2002; 2002US-0410953P.
PR   17-SEP-2002; 2002US-0410957P.
PR   17-SEP-2002; 2002US-0410958P.
PR   17-SEP-2002; 2002US-0410959P.
PR   17-SEP-2002; 2002US-0410960P.
PR   17-SEP-2002; 2002US-0410961P.
PR   17-SEP-2002; 2002US-0410962P.
PR   17-SEP-2002; 2002US-0411019P.
PR   17-SEP-2002; 2002US-0411022P.
PR   17-SEP-2002; 2002US-0411023P.
PR   17-SEP-2002; 2002US-0411024P.
PR   17-SEP-2002; 2002US-0411032P.
PR   17-SEP-2002; 2002US-0411035P.
PR   17-SEP-2002; 2002US-0411037P.
PR   17-SEP-2002; 2002US-0411041P.
PR   17-SEP-2002; 2002US-0411045P.
PR   17-SEP-2002; 2002US-0411046P.
PR   17-SEP-2002; 2002US-0411048P.
PR   17-SEP-2002; 2002US-0411052P.
PR   17-SEP-2002; 2002US-0411055P.
PR   17-SEP-2002; 2002US-0411073P.
PR   17-SEP-2002; 2002US-0411082P.
PR   17-SEP-2002; 2002US-0411101P.
PR   17-SEP-2002; 2002US-0411111P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Beaurang PA,  Behrens D;
PI   Halenbeck RF,  Huang MM,  Kothakota S,  Haishan L,  Linnemann T;
PI   Pierce K,  Wang Y,  Wong JGP,  Wu G,  Zhang H;
XX
DR   WPI; 2004-348438/32.
XX
PT   New nucleic acid molecule for diagnosing, preventing or treating diseases
PT   such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
PT   genetic, bacterial and viral diseases.
XX
PS   Claim 1; SEQ ID NO 822; 428pp; English.
XX
CC   The present invention relates to an isolated nucleic acid molecule
CC   encoding a polypeptide which is believed to be cytostatic,
CC   antiinflammatory, immunosuppressive, antibacterial and virucidal. The
CC   composition and methods are useful for diagnosing, preventing and
CC   treating diseases such as proliferative (e.g. cancer), inflammatory,
CC   immune, metabolic, genetic, bacterial and viral diseases. The present
CC   sequence represents a human secreted protein encoding sequence. The
CC   present sequence is available on WIPOWEB and is not in the specification.
XX
SQ   Sequence 1470 BP; 215 A; 578 C; 486 G; 191 T; 0 U; 0 Other;

  Query Match             23.4%;  Score 69;  DB 12;  Length 1470;
  Best Local Similarity   55.7%;  Pred. No. 0.1;
  Matches  132;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         45 GCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGC 104
              || | | ||||  |||  || ||||| | ||  |   ||| |  ||||  || || ||||
Db        783 GCCGCCCGCGCCGCGCCGCGACCCGGGGCCGGGGACCGCGCGGGGGCGGCCGGGGGGCGC 724

Qy        105 GGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCG 164
              | || ||||||||     |    || |  ||  ||   |||  | | |  || |  || |
Db        723 CGGGGGCCGCGGCCGGGGCGCGCCGGGACCGGGGACGCGCGGGCCGGGGCCGGGTCCGGG 664

Qy        165 GGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCC 224
              | | ||||| ||    |   || | |   | |  |||||  |||  | ||  ||| |||
Db        663 CGAGGGCGCGCGGGCGCCGCGGGCCGGGCGGGCGCCGGGGCGGGGGCGGGGACCGGGGCC 604

Qy        225 GGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCC 281
              || || | |  |  |  ||  |    | ||   |  |  |||  || || |||||||
Db        603 GGGCAAGGCTCCCGTCAGGGGCCGGCGGGCGGGCGCGCGCGGGGCGCCGGGGGGCC 547


RESULT 6
ADZ51725
ID   ADZ51725 standard; DNA; 138203 BP.
XX
AC   ADZ51725;
XX
DT   14-JUL-2005  (first entry)
XX
DE   FR-008 polyketide gene cluster, SEQ ID 1.
XX
KW   Antifungal; antiparasitic; polyketide; gene; ds.
XX
OS   Streptomyces sp.; FR-008.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(574. .1950)
FT                   /*tag=  a
FT                   /product= "FAD-dependent monooxygenase, FscO, SEQ ID 20"
FT   CDS             complement(2264. .3037)
FT                   /*tag=  b
FT                   /product= "ADC lyase, PabC, SEQ ID 22"
FT   CDS             complement(3150. .3818)
FT                   /*tag=  c
FT                   /product= "Regulator protein, FscRI, SEQ ID 10"
FT   CDS             complement(4377. .7205)
FT                   /*tag=  d
```

```
FT                   /product= "Regulator protein, FscRII, SEQ ID 11"
FT   CDS             complement(7210. .10320)
FT                   /*tag= e
FT                   /product= "Regulator protein, FscRIII, SEQ ID 12"
FT   CDS             complement(10298. .13315)
FT                   /*tag= f
FT                   /product= "Regulator protein, FscRIV, SEQ ID 13"
FT   CDS             13522. .14898
FT                   /*tag= g
FT                   /product= "Glycosyltransferase, FscMI, SEQ ID 17"
FT   CDS             14953. .16011
FT                   /*tag= h
FT                   /product= "GDP-ketosugar aminotransferase, FscMII, SEQ ID
FT                   18"
FT   CDS             16106. .17287
FT                   /*tag= i
FT                   /product= "Cytochrome P-450 monooxygenase, FscP, SEQ ID
FT                   14"
FT   CDS             17334. .17528
FT                   /*tag= j
FT                   /product= "Ferredoxin protein, FscFE, SEQ ID 15"
FT   CDS             17556. .18413
FT                   /*tag= k
FT                   /product= "Thioesterase, FscTE, SEQ ID 16"
FT   CDS             18610. .20781
FT                   /*tag= l
FT                   /product= "ADC synthase, PabAB, SEQ ID 21"
FT   CDS             20927. .26158
FT                   /*tag= m
FT                   /product= "PKS FscA, SEQ ID 2"
FT   CDS             26333. .27340
FT                   /*tag= n
FT                   /product= "ABC transporter, FscTI, SEQ ID 8"
FT   CDS             27561. .28280
FT                   /*tag= o
FT                   /product= "ABC transporter, FscTII, SEQ ID 9"
FT   CDS             complement(28983. .60860)
FT                   /*tag= p
FT                   /product= "PKS FscC, SEQ ID 3"
FT   CDS             complement(60962. .77587)
FT                   /*tag= q
FT                   /product= "PKS FscB, SEQ ID 4"
FT   CDS             complement(77983. .84132)
FT                   /*tag= r
FT                   /product= "PKS FscF, SEQ ID 5"
FT   CDS             complement(84170. .107485)
FT                   /*tag= s
FT                   /product= "PKS FscE, SEQ ID 6"
FT   CDS             complement(107496. .136148)
FT                   /*tag= t
FT                   /product= "PKS FscD, SEQ ID 7"
FT   CDS             complement(136558. .137766)
FT                   /*tag= u
FT                   /product= "GDP-mannose-4,6-dehydratase, FscMIII, SEQ ID
FT                   19"
XX
PN   US2005089884-A1.
XX
PD   28-APR-2005.
XX
PF   06-APR-2004; 2004US-00819386.
XX
PR   23-OCT-2003; 2003KR-00074035.
XX
PA   (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
PA   (UYSH-) UNIV SHANGHAI JIAOTONG.
XX
PI   Lee SY,  Deng Z,  Chen S,  Jeong KJ,  Zhou X;
XX
DR   WPI; 2005-322033/33.
DR   P-PSDB; ADZ51726, ADZ51727, ADZ51728, ADZ51729, ADZ51730, ADZ51731,
DR   ADZ51732, ADZ51733, ADZ51734, ADZ51735, ADZ51736, ADZ51737, ADZ51738,
DR   ADZ51739, ADZ51740, ADZ51741, ADZ51742, ADZ51743, ADZ51744, ADZ51745,
DR   ADZ51746.
XX
PT   New gene cluster for the biosynthesis of FR-008 polyketides, useful for
PT   increasing the productivity of FR-008 polyketides or to produce new FR-
PT   008 polyketide variants.
XX
PS   Claim 1; SEQ ID NO 1; 22pp; English.
XX
CC   The present invention relates to a gene cluster (ADZ51725) for the
CC   biosynthesis of FR-008 polyketides derived from Streptomyces sp. FR-008.
CC   FR-008 polyketide, a hetaene macrolide having aglycone containing 4-
CC   aminoacetophenone, has antifungal activity and also high toxicity against
CC   mosquito larvae, and so is highly useful in agricultural and medical
CC   fields. The gene cluster comprises the following genes: fscA, fscC, fscB,
CC   fscF, fscE and fscD, which code for modular polyketide synthase (PKS);
CC   fscTI and fscTII, which code for ABC transporter proteins; fscRI, fscRII,
CC   fscRIII and fscRIV genes which code for regulator proteins; fscP which
CC   codes for cytochrome P450 monooxygenase; fscFE which codes for ferredoxin
CC   protein; fscTE which codes for thioesterase; fscMI which codes for
CC   glycosyltransferase; fscMII which codes for GDP-ketosugar
CC   aminotransferase; fscMIII which codes for GDP-mannose-4,6-dehydratase;
CC   fscO which codes for FAD-dependent monooxygenase; pabAB which codes for 4
CC   -amino-4-deoxychorismate (ADC) synthase and pabC which codes for ADC
CC   lyase. The genes of the gene cluster of the invention can be used to
CC   develop recombinant microorganisms capable of producing FR-008
CC   polyketides, also to increase the productivity of the existing FR-008
CC   polyketides or to produce new FR-008 polyketide variants, by its
CC   modification.
XX
SQ   Sequence 138203 BP; 18153 A; 48202 C; 55342 G; 16506 T; 0 U; 0 Other;

  Query Match             22.4%;  Score 66;  DB 14;  Length 138203;
  Best Local Similarity   51.7%;  Pred. No. 0.08;
  Matches  150;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              ||||||    |     ||   |  | |   |  || ||||||  |   | ||||| || ||
Db     126896 CGAGGGTCTTGGCGGCCATGGCGGCGGCGACCTGGTCGAGCGGGGTGTCGGCGAGCGGCG 126955

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
               | |  | || |||||  |||||||| |    ||  ||| | |||| ||   ||| ||
Db     126956 CGGCCTGGCCGACGCCGGCGGCGTGCACGACGCCGGTGAGGGGGGCGTCCTCGGGCACCG 127015

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
                 || |||| |||       | |||    |||  |||| ||||     |  ||
Db     127016 CGGCGAGGACGGCGGCGAGGGCGTCGCGGTCGGCGCTGTCGCAGGCGGTGAGGGTGACGC 127075

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              ||  |||  | |||     |||| |  |  ||||||||||     | ||  |  |
Db     127076 GGGCGCCCAGTTCCTCCAGTTCGGCGCGGAGTTCGGCGGCGCCGGGGGCGTCGGGGCCGC 127135

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGC 291
              |||   |       | |||   |   || ||  | || |||||  | ||
Db     127136 GGCGGCCGGTGAGGACGAGGTGGGCGGCGCCGGAGCGGGCGAGCCAGCGC 127185


RESULT 7
ACL67140/c
ID   ACL67140 standard; DNA; 1338 BP.
XX
AC   ACL67140;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus gene sequence, seq id 3603.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression;
KW   gene; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
```

PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.
XX
PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 2; SEQ ID NO 3603; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC set of about 7842 genes or partial genes from the genome of the bacterium
CC Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from USPTO
XX
SQ Sequence 1338 BP; 177 A; 454 C; 514 G; 193 T; 0 U; 0 Other;

Query Match 21.8%; Score 64.4; DB 14; Length 1338;
Best Local Similarity 51.8%; Pred. No. 0.45;
Matches 146; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              ||||| | || | ||  || |||  | | |||| ||   || |||| | | | | | ||
Db       1269 GGCGGCCAGCTCTTCCCGGGCCTGGGGCAGGCCCAGGTTGCCGGCGCCACCCAGGTGGGT 1210

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
                | |    | ||| |  ||   | ||          |  |   |    | |||  | |
Db       1209 GGCCGTCAGGGCGCCGCGGTCTGGACGGAACGTTCCTTCCACCAGCTCGCGGCCCACCTG 1150

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              || |  ||||||      ||| | | |  |  | |  | || |  || ||| ||||| |
Db       1149 CTTGTAGGCGTCACGGAAGGCCTGTCCCTGGGCCACCAGCGCATAGGCATGGTGCGCCGC 1090

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              | |||||| ||| |||  |||  ||| || |||| |  |  |||     |     |||  |
Db       1089 GTACAGCGTGTCGTCGCTGGCACGCGCCGCCGCGTCCGCCTTCACCTGGAGCGCGGGCAC 1030

Qy        246 CACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAG 287
              || | | |    || | || |    ||  | | |  |  |||
Db       1029 CAACCGGGTGAGCACGTCCAGCAGCTCCCGCGTGGAGCCCAG 988
```

RESULT 8
ACL64710/c
ID ACL64710 standard; DNA; 20113 BP.
XX
AC ACL64710;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus DNA fragment, seq id 1173.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.

XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.
XX
PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 1; SEQ ID NO 1173; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC set of 1849 contig and singleton sequences comprising coding sequences,
CC DNA replication elements, promoters and other regulatory elements from
CC the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO
XX
SQ Sequence 20113 BP; 3209 A; 7129 C; 6944 G; 2831 T; 0 U; 0 Other;

Query Match 21.8%; Score 64.4; DB 14; Length 20113;
Best Local Similarity 51.8%; Pred. No. 0.22;
Matches 146; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              ||||| | || | ||  || |||  | | |||| ||   || |||| | | | | | ||
Db      19270 GGCGGCCAGCTCTTCCCGGGCCTGGGGCAGGCCCAGGTTGCCGGCGCCACCCAGGTGGGT 19211

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
                | |    | ||| |  ||   | ||          |  |   |    | |||  | |
Db      19210 GGCCGTCAGGGCGCCGCGGTCTGGACGGAACGTTCCTTCCACCAGCTCGCGGCCCACCTG 19151

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              || |  ||||||      ||| | | |  |  | |  | || |  || ||| ||||| |
Db      19150 CTTGTAGGCGTCACGGAAGGCCTGTCCCTGGGCCACCAGCGCATAGGCATGGTGCGCCGC 19091

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              | |||||| ||| |||  |||  ||| || |||| |  |  |||     |     |||  |
Db      19090 GTACAGCGTGTCGTCGCTGGCACGCGCCGCCGCGTCCGCCTTCACCTGGAGCGCGGGCAC 19031

Qy        246 CACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAG 287
              || | | |    || | || |    ||  | | |  |  |||
Db      19030 CAACCGGGTGAGCACGTCCAGCAGCTCCCGCGTGGAGCCCAG 18989
```

RESULT 9
AAL61180
ID AAL61180 standard; DNA; 1047 BP.
XX
AC AAL61180;
XX
DT 22-SEP-2003 (first entry)
XX
DE Actinosynnema pretiosum methyltransferase gene #1.
XX
KW Maytansinoid; ansamitocin; antitumour; methyltransferase; enzyme; gene;
KW ds.
XX

OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
DR WPI; 2003-493374/46.
XX
PT Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS Disclosure; Page 72; 160pp; English.
XX
CC The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumour agent. The present sequence is A. pretiosum
CC ansamitocin gene cluster I methyltransferase gene
XX
SQ Sequence 1047 BP; 106 A; 414 C; 403 G; 124 T; 0 U; 0 Other;

Query Match 21.6%; Score 63.8; DB 8; Length 1047;
Best Local Similarity 51.6%; Pred. No. 0.59;
Matches 146; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy         13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
              | | || || |||   |       |||    | ||  || || | || | ||  |||
Db        458 TCCACGACGCGGTGGCCACCGGCCCCGGCGACCCGTTCGACGAGGCGATCCGCTCCGCGC 517

Qy         73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
              ||| |  |  || |  |||||| | ||    || ||||    |||      |||    |
Db        518 TCGACCTGACCGGGGCCGCGCGGGTGGTCGACGTCGGCGGCGGGCGGGGCAGGCTGCTCG 577

Qy        133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
               ||     ||||||  |   | || |||| |||||||||   | || |  | ||    |
Db        578 CCGCGCTGCTGCGCGAGCACCCCGTGCTCTCGGGCGTGCTGTTCGACCTGCCCGAGGCGG 637

Qy        193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
              |     || ||       || |||||||| ||  | || |       ||||   |   |
Db        638 TCCGGGACGCGGTTCCCGAGCTGCGCGGCGACGACCGGGTCCAGATCAGGGCGGGCGACC 697

Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
                |||  |||||||  || || || || ||   | |||| |||
Db        698 TGTTCGCGGACCGGGTCGGCGCGGCGGACGTCTACGTGCTGCG 740
```

RESULT 10
AAL61224
ID AAL61224 standard; DNA; 82746 BP.
XX
AC AAL61224;
XX
DT 22-SEP-2003 (first entry)
XX
DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.
XX
KW Maytansinoid; ansamitocin; antitumour; gene; ds.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
DR WPI; 2003-493374/46.
XX
PT Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS Claim 7; Page 105-152; 160pp; English.
XX
CC The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumour agent. The present sequence is A. pretiosum
CC ansamitocin ansamitocin biosynthetic gene cluster I
XX
SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;

Query Match 21.6%; Score 63.8; DB 8; Length 82746;
Best Local Similarity 51.6%; Pred. No. 0.19;
Matches 146; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy         13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
              | | || || |||   |       |||    | ||  || || | || | ||  |||
Db      76046 TCCACGACGCGGTGGCCACCGGCCCCGGCGACCCGTTCGACGAGGCGATCCGCTCCGCGC 76105

Qy         73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
              ||| |  |  || |  |||||| | ||    || ||||    |||      |||    |
Db      76106 TCGACCTGACCGGGGCCGCGCGGGTGGTCGACGTCGGCGGCGGGCGGGGCAGGCTGCTCG 76165

Qy        133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
               ||     ||||||  |   | || |||| |||||||||   | || |  | ||    |
Db      76166 CCGCGCTGCTGCGCGAGCACCCCGTGCTCTCGGGCGTGCTGTTCGACCTGCCCGAGGCGG 76225

Qy        193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
              |     || ||       || |||||||| ||  | || |       ||||   |   |
Db      76226 TCCGGGACGCGGTTCCCGAGCTGCGCGGCGACGACCGGGTCCAGATCAGGGCGGGCGACC 76285

Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
                |||  |||||||  || || || || ||   | |||| |||
Db      76286 TGTTCGCGGACCGGGTCGGCGCGGCGGACGTCTACGTGCTGCG 76328
```

RESULT 11
ABI99206
ID ABI99206 standard; cDNA; 3959 BP.
XX
AC ABI99206;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:8.
XX
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS Mus musculus.
XX
PN WO200188188-A2.
XX
PD 22-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-JP004192.
XX
PR 18-MAY-2000; 2000JP-00145977.
XX
PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

```
XX
PI   Ishikawa K,  Asai S,  Takahashi Y,  Nagata T,  Ishii Y;
XX
DR   WPI; 2002-034733/04.
DR   P-PSDB; ABB57023.
XX
PT   Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT   expression levels of particular genes defined in the specification or by
PT   determining the expression profile of a gene group comprising these
PT   genes.
XX
PS   Claim 2; Page 64-72; 2690pp; English.
XX
CC   The present invention describes a method for examining ischaemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (I) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (I). The method
CC   is useful for examining the ischaemic condition (e.g. compressive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding the
CC   protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The expression
CC   levels or expression profiles produced by these genes are used as an
CC   indicator when screening for ischaemic condition-improving drugs or
CC   therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
CC   primers for a mouse ischaemic condition related sequence, which are used
CC   in the exemplification of the present invention
XX
SQ   Sequence 3959 BP; 1104 A; 1122 C; 1197 G; 536 T; 0 U; 0 Other;

  Query Match             21.6%;  Score 63.6;  DB 6;  Length 3959;
  Best Local Similarity   56.8%;  Pred. No. 0.44;
  Matches  117;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy         88 AGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCC 147
              ||| |||||||  ||||||||  |  || || ||    |  | | |||   | |||||||
Db        593 AGGTGCGCGAGATGCGCGGCGCCGTGCTGCGCCTCGGGGCGGCGCGCGGGCAGCTGCGCC 652

Qy        148 CGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCG 207
               || ||| | || |  |   | |  | | |  | | || | |||||  |   || ||  |
Db        653 TGGAGCAGGAGCACCTGCTGGAGGACATCGCTCACGTCCGCCAGCGGCTGGACGAGGAGG 712

Qy        208 GGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGA 267
                |||| ||| |  |  |   |||| |  |   | ||||   |||||||  | |   |
Db        713 CCCGGCAGCGTGAGGAGGCGGAGGCGGCGGCGCGCGCCCTGGCGCGCTTCGCGCAGGAGG 772

Qy        268 CCGTCGTGGGGGCCGAGCAGGTGCAG 293
              | |  | || |  || | || |||||
Db        773 CGGAAGCGGCGCGCGTGGAGCTGCAG 798


RESULT 12
ABX04971/c
ID   ABX04971 standard; DNA; 103599 BP.
XX
AC   ABX04971;
XX
DT   16-JAN-2003  (first entry)
XX
DE   S. cinnamonensis monensin type I polyketide synthase gene cluster.
XX
KW   Monensin; gene; cluster; polyketide synthase; antibiotic; ds;
KW   antihelminthic; insecticide; immunosuppressant; antifungal;
KW   antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;
KW   mon RI; mon RII; mon T; mon AIX; mon AX.
XX
OS   Streptomyces cinnamonensis.
XX
PN   WO200168867-A1.
XX
PD   20-SEP-2001.
XX
PF   30-MAY-2000; 2000WO-GB002072.
XX
PR   28-MAY-1999;  99GB-00012563.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
XX
PI   Leadlay PF,  Staunton J,  Oliynyk M;
XX
DR   WPI; 2001-611393/70.
DR   P-PSDB; ABG99854, ABG99855, ABG99856, ABG99857, ABG99858, ABG99859,
DR   ABG99860, ABG99861, ABG99862, ABG99863, ABG99864, ABG99865, ABG99866,
DR   ABG99867, ABG99868, ABG99869, ABG99870, ABG99871, ABG99872, ABG99873,
DR   ABG99874, ABG99875, ABG99876, ABG99877, ABG99878, ABG99879, ABG99880,
DR   ABG99881, ABG99882, ABG99883, ABG99884, ABG99885, ABG99886, ABG99887.
XX
PT   New DNA sequence encoding polyketide synthase, useful for the production
PT   of polyketides such as antibiotic monensin.
XX
PS   Claim 1; Page 116-195; 212pp; English.
XX
CC   The invention relates to a DNA sequence which is a fully defined sequence
CC   of 103551 base pairs appearing as ABX04971, or its variant, that it is
CC   not a sequence encoding all or part amino acids 1-920 encoded by mon AI
CC   as given in the specification. The DNA is the S. cinnamonensis polyketide
CC   antibiotic monensin biosynthetic gene cluster. Also included are a
CC   recombinant cloning or expression vector comprising the gene cluster, a
CC   transformed host cell which has been transformed to contain the gene
CC   cluster (and is capable of expressing a corresponding polypeptide), a
CC   hybridization probe derived from the gene cluster (for identification and
CC   isolation of the same or analogous gene cluster, e.g. one which binds
CC   specifically to a region of the monensin gene cluster selected from mon
CC   BI, mon BII, mon CI, mon CII, mon H, Mon RI, mon RII, mon T, mon AIX and
CC   mon AX), the use of the mon RI gene or variant and a monensin promoter to
CC   control expression of a heterologous gene in Streptomyces cinnamonensis,
CC   a polypeptide encoded by a portion of the monensin gene cluster
CC   (preferably comprising mon BI, mon BII, mon AIX or mon AX or their
CC   mutants, alleles or variants), an epoxidase enzyme encoded by mon CI, a
CC   cyclase enzyme encoded by mon CII, producing S. cinnamonensis capable of
CC   enhanced levels of production of monensin comprising engineering it to
CC   overexpress the mon RI gene, S. cinnamonensis containing multiple copies
CC   of the mon RI gene and/or its variants, expressing a gene heterologous to
CC   S. cinnamonensis comprising transforming S. cinnamonensis with DNA
CC   encoding a heterologous gene and expressing the gene under control of the
CC   activator gene mon RI or actII/orf4 and 13-propyl erythromycin A. The
CC   processes and materials (enzyme systems, nucleic acids and vectors) are
CC   useful for preparing polyketides by recombinant synthesis. The
CC   polyketides are useful as insecticides, antibiotics, anthelminthics,
CC   antifungals, antibacterials or other pharmaceuticals. In particular the
CC   gene is useful for the production of monensin, an antibiotic polyether
CC   polyketide. The present sequence represents the monensin gene cluster
XX
SQ   Sequence 103599 BP; 13980 A; 37023 C; 37799 G; 14795 T; 0 U; 2 Other;

  Query Match             21.5%;  Score 63.4;  DB 4;  Length 103599;
  Best Local Similarity   52.5%;  Pred. No. 0.2;
  Matches  139;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              | | | | | ||| ||  | || | | |  |  | ||  |  || |  |  || |   ||
Db      41831 GCCCGCCGGGGCGACGCTGCCCGCGGCCGCGGTGCGCCGGGCGGTGCGGGCCCCCGGGGT 41772

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
              | |  |  || |  ||||| ||||| ||||| |  | |  | |||  || || |
Db      41771 CGCACCCGCGGCCAGGGCGCGCAGGAGCGCGGGCCGGGACGGGGCGGGCTGCGCGGCGAC 41712

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
                || | |||||||  | | ||  |  |||| ||  |||| |||| ||| ||   || ||
Db      41711 GGCGCGCGCGTCGAGGTCCACCGCGACCACGTGCGGGCGGCCGTGGGCGCGGGCGGCGTC 41652

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              |   || ||      |     |||  | | ||| || |         ||  |  ||| | |
Db      41651 GAGGAGGGCCAAGCCACCTTCGGTGGTCAGCGGGGTGAACCCGGTGCGCTTCATGCGTGC 41592

Qy        246 CACCTGCGCTTCCAGGACCGGACCG 270
```

```
                ||  | ||  ||   ||   | |||
Db      41591 CAGGTCCGTCTCGCCGAGGTGTCCG 41567


RESULT 13
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   15-SEP-2003  (revised)
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             3. .992
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= "transposase"
FT                   /note= "not part of the PKS"
FT   CDS             989. .1501
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= "transposase"
FT                   /note= "not part of the PKS"
FT   CDS             1998. .6263
FT                   /*tag=  c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031. .3548
FT                   /*tag=  d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        2031. .3548
FT                   /*tag=  i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031. .3548
FT                   /*tag=  j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        3621. .4661
FT                   /*tag=  e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917. .5810
FT                   /*tag=  f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the thiazole
FT                   moiety"
FT   misc_RNA        5856. .6155
FT                   /*tag=  g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260. .10493
FT                   /*tag=  h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        6861. .6887
FT                   /*tag=  k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861. .6887
FT                   /*tag=  l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358. .7366
FT                   /*tag=  m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898. .7921
FT                   /*tag=  n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898. .7921
FT                   /*tag=  o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261. .8308
FT                   /*tag=  p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411. .8422
FT                   /*tag=  q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861. .8905
FT                   /*tag=  r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966. .8983
FT                   /*tag=  s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090. .9179
FT                   /*tag=  t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183. .9992
FT                   /*tag=  u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121. .10138
FT                   /*tag=  v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261. .10306
FT                   /*tag=  w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639. .16137
FT                   /*tag=  x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654. .12033
FT                   /*tag=  y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250. .13287
FT                   /*tag=  z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327. .13899
FT                   /*tag=  aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962. .15756
FT                   /*tag=  ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763. .16008
FT                   /*tag=  ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134. .37907
FT                   /*tag=  ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425. .17606
FT                   /*tag=  ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817. .18857
FT                   /*tag=  af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581. .20396
FT                   /*tag=  ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424. .20642
FT                   /*tag=  ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706. .22082
```

```
FT                   /*tag=  ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296. .23336
FT                   /*tag=  aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069. .24647
FT                   /*tag=  ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867. .25151
FT                   /*tag=  al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203. .26576
FT                   /*tag=  am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793. .27883
FT                   /*tag=  an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966. .28574
FT                   /*tag=  ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433. .30287
FT                   /*tag=  ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321. .30869
FT                   /*tag=  aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077. .31373
FT                   /*tag=  ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440. .32807
FT                   /*tag=  as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018. .34067
FT                   /*tag=  at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107. .34676
FT                   /*tag=  au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760. .36641
FT                   /*tag=  av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705. .37256
FT                   /*tag=  aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470. .37769
FT                   /*tag=  ax
FT                   /note= "encodes ACP6"
FT   CDS             37912. .49308
FT                   /*tag=  ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014. .39375
FT                   /*tag=  az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589. .40626
FT                   /*tag=  ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341. .41922
FT                   /*tag=  bb
FT                   /note= "encodes KR7"

  Query Match             21.2%;  Score 62.4;  DB 3;  Length 71989;
  Best Local Similarity   51.4%;  Pred. No. 0.31;
  Matches  144;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         14 GCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGT 73
              |||||||| || |  |    |||| |||| || |  ||||      || || ||  |||
Db      65289 GCGCGTCGCGGCCGGCGTCGCGGCGGAGCTCGTGCACGGCCTCGGCCTGCGCCCGCGCGG 65348

Qy         74 CGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGG 133
                | ||| ||   | ||   || |  |||||    ||  |  ||          | |
Db      65349 GATCACGGACGCCGACGCCGCCGCGCTGCGCGCGCTCCGCCCCCCGGACGGCGAGGGGCA 65408

Qy        134 CGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCG 193
              || ||   |   || ||||| ||||| ||    || ||||| || | ||  |||
Db      65409 CGGCGCCTTCGTCCGGGCGCTCGCGCGCGATCTCGCGCGCGCGGGGGGCGCCGGCGTCGC 65468

Qy        194 CGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCG 253
              |||| |||| | ||| | || || | || |          ||   | |    | ||
Db      65469 CGTCGTCGGCGACGGCCAGCCGCCCATCGTCCACGCCCTCGGGCACGTCATCAACGCCGC 65528

Qy        254 CTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
                |||    |||| | | |  |  || ||| |  |||| |
Db      65529 GCTCCGCAGCCGGGCGGCCTGGATGGTCGATCCTGTGCTG 65568


RESULT 14
ADX26149
ID   ADX26149 standard; DNA; 3403 BP.
XX
AC   ADX26149;
XX
DT   05-MAY-2005  (first entry)
XX
DE   Novel cell pain response detection method-related rat gene SeqID495.
XX
KW   pain; animal disease model; expression; analgesic; antiaddictive;
KW   nootropic; anticonvulsant; vasotropic; neuroprotective; tranquilizer;
KW   antiasthmatic; antirheumatic; antiarthritic; osteopathic;
KW   ophthalmological; antiinflammatory; antipruritic; dermatological;
KW   antiulcer; gastrointestinal-Gen.; nephrotropic; gynecological;
KW   hepatotropic; antiparkinsonian; neuroleptic; laxative; gene therapy;
KW   neuropathic pain; Alzheimers disease; Parkinsons disease;
KW   motor neurone disease; Huntingtons disease; schizophrenia; gene; ds.
XX
OS   Rattus sp.
XX
PN   WO2005014849-A2.
XX
PD   17-FEB-2005.
XX
PF   06-JUL-2004; 2004WO-US023166.
XX
PR   03-JUL-2003; 2003US-0485101P.
XX
PA   (EURO-) EUROCELTIQUE SA.
XX
PI   Tong J,  Jin G,  Ji R,  Xu Y,  Chiang LW,  Lavery DJ;
XX
DR   WPI; 2005-163258/17.
XX
PT   Detecting pain responses in a cell, useful in identifying potential
PT   therapeutic and diagnostic candidates for treating pain, by identifying
PT   genes that are differentially expressed in a model of neuropathic pain.
XX
PS   Example 1; SEQ ID NO 495; 173pp; English.
XX
CC   This invention relates to a novel method of detecting a pain response in
CC   a cell which comprises determining the expression level in a test cell of
CC   at least one nucleic acid molecule and comparing the expression level to
CC   a level in an animal model of pain, where similar or identical expression
CC   levels indicate a pain response in the test cell. The invention may be
CC   useful for the development of compounds with an analgesic, antiaddictive,
CC   nootropic, anticonvulsant, vasotropic, neuroprotective, tranquilizer,
CC   antiasthmatic, antirheumatic, antiarthritic, osteopathic,
CC   ophthalmological, antiinflammatory, antipruritic, dermatological,
CC   antiulcer, gastrointestinal-Gen., nephrotropic, gynecological,
CC   hepatotropic, antiparkinsonian, neuroleptic or laxative activity whilst
CC   the disclosed sequences may prove useful for gene therapy. The methods
CC   and compositions of the present invention are useful for identifying
CC   agonists and antagonists for the gene or gene products as potential
CC   therapeutic and diagnostic candidates for treating pain, including
CC   neuropathic pain, nociceptive pain, chronic pain, inflammatory pain, pain
CC   associated with cancer, and pain associated with rheumatic disease, and
CC   also for addiction, seizure, stroke, ischemia, a neurodegenerative
CC   disorder, anxiety, depression, headache, asthma, rheumatic disease,
```

CC osteoarthritis, retinopathy, inflammatory eye disorders, pruritus, ulcer,
CC gastric lesions, uncontrollable urination, an inflammatory or unstable
CC bladder disorder, inflammatory bowel disease, irritable bowel syndrome
CC (IBS), irritable bowel disease (IBD), gastroesophageal reflux disease
CC (GERD), functional dyspepsia, functional chest pain of presumed
CC esophageal origin, functional dysphagia, non-cardiac chest pain,
CC symptomatic gastroesophageal disease, gastritis, aerophagia, functional
CC constipation, functional diarrhea, burbulence, chronic functional
CC abdominal pain, recurrent abdominal pain (RAP), function abdominal
CC bloating, functional biliary pain, functional incontinence, functional
CC ano-rectal disorder, cholecystalgia, interstitial cystitis, dysmenorrhea,
CC or dyspareunia. They can also be used diagnosing or treating Alzheimer's
CC disease, Parkinson's disease, amyotrophic lateral sclerosis, Huntington's
CC disease and schizophrenia. The present sequence is that of a rat gene
CC which demonstrated homology to a human gene whose expression level
CC analysed in the method of the invention.
XX
SQ Sequence 3403 BP; 936 A; 963 C; 1082 G; 422 T; 0 U; 0 Other;

Query Match 21.0%; Score 62; DB 14; Length 3403;
Best Local Similarity 56.3%; Pred. No. 0.77;
Matches 116; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy        88 AGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCC 147
             ||| |||||||  |||||| | |  || || ||    |   | | |||   |  ||||||
Db       497 AGGTGCGCGAGATGCGCGGGGCCGTGCTGCGCCTGGGGGCGGCGCGCGGGCACGTGCGCC 556

Qy       148 CGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCG 207
             || ||| | || |  |   | |  | | | || | || | |||||  |   || ||  |
Db       557 TGGAGCAGGAGCACTTGCTGGAGGACATCGCGCACGTCCGCCAGCGGCTGGACGAGGAGG 616

Qy       208 GGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGA 267
             ||||  |||||  |  |   |||| |  |   | ||||    ||| |||  | |   |
Db       617 CCCGGCAGCGCGAGGAAGCGGAGGCGGCGGCCCGCGCCCTAGCGCGGTTCGCGCAGGAGG 676

Qy       268 CCGTCGTGGGGGCCGAGCAGGTGCAG 293
             | |  | || |  || | || |||||
Db       677 CGGAAGCGGCGCGCGTGGAGCTGCAG 702
```

RESULT 15
ACA26420/c
ID ACA26420 standard; DNA; 2184 BP.
XX
AC ACA26420;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #8077.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Burkholderia mallei.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU22550.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 14290; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2184 BP; 262 A; 801 C; 831 G; 290 T; 0 U; 0 Other;

Query Match 20.9%; Score 61.8; DB 8; Length 2184;
Best Local Similarity 52.9%; Pred. No. 0.93;
Matches 157; Conservative 0; Mismatches 137; Indels 3; Gaps 1;

```
Qy         1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
             ||||| |||  |||||||||  ||| || | |  ||          | |  ||| ||  |
Db       683 GCGAGCGCGAGCTGCGCGTCGGCGTCGTCGGGCGCGCGCGCGAGCAGCGGCGCGAGCGCC 624

Qy        61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
               | ||| ||||  |||  | |||  |||||||||    || ||  |  |||
Db       623 GCGCGCGCGGTGTCGTAGCGCCCCTGCGCGCGCGCGAGCTCGGCCGCCTCGAGCGCATAG 564

Qy       121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
             |  || |||     | |   | |   |  |||  || || |||| |||  ||| | |
Db       563 CGGCGATCGGCCCTGCCTGCCGGGCTCGCGCGATCGAGCGCGCGCGCGGCCGCCTCGTAA 504

Qy       181 GCCTCG---GACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCG 237
               | |||      |||||||| ||||   || || ||            | | |  | |||
Db       503 TCGTCGCGATCGAGCGCGTCATCGGTTTCGCGCAGCACGAGCCGCAGCGCCTGATCCTCG 444

Qy       238 ATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGC 294
             | | |  |   |||||  |   || |||  || ||  | |||  |||  ||  |||
Db       443 AGGCGCGCGCTTTGCGCGGCGTCGAGCGGCGCGCCGCCGCGGCGCAGCGTGTCGAGC 387
```

Search completed: April 27, 2006, 08:00:20
Job time : 145.252 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:21 ; Search time 60.863 Seconds
(without alignments)
8615.758 Million cell updates/sec

Title: US-10-729-802-1_COPY_77000_77294
Perfect score: 295
Sequence: 1 gcgagggcggtctgcgcgtc.....ggggccgagcaggtgcagcg 295

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	64.4	21.8	1338	3	US-09-902-540-3603	Sequence 3603, Ap
c	2	64.4	21.8	20113	3	US-09-902-540-1173	Sequence 1173, Ap
	3	62.4	21.2	71989	3	US-09-443-501A-2	Sequence 2, Appli
	4	62.4	21.2	71989	3	US-09-724-889A-2	Sequence 2, Appli
	5	62.4	21.2	71989	3	US-09-724-876-2	Sequence 2, Appli
c	6	61.2	20.7	1683	3	US-09-758-759-148	Sequence 148, App
c	7	61.2	20.7	58857	3	US-09-477-962-1	Sequence 1, Appli
c	8	61.2	20.7	109519	3	US-09-758-759-1	Sequence 1, Appli
c	9	60.8	20.6	4257	2	US-08-690-473-1	Sequence 1, Appli
c	10	60.8	20.6	4257	3	US-09-259-821A-1	Sequence 1, Appli
c	11	60.8	20.6	4257	3	US-08-843-659-1	Sequence 1, Appli
c	12	60.8	20.6	4257	3	US-09-825-288A-1	Sequence 1, Appli
	13	60.8	20.6	12001	2	US-08-458-568A-11	Sequence 11, Appl
	14	59.8	20.3	1162	2	US-08-726-306A-52	Sequence 52, Appl
c	15	59.6	20.2	3957	3	US-10-237-551-193	Sequence 193, App
	16	59.6	20.2	154746	3	US-09-827-688-8	Sequence 8, Appli
c	17	59.6	20.2	154746	3	US-09-827-688-8	Sequence 8, Appli
	18	58.2	19.7	1449	3	US-09-902-540-7506	Sequence 7506, Ap
c	19	58.2	19.7	5618	3	US-09-902-540-728	Sequence 728, App
c	20	57.2	19.4	1500	3	US-09-593-711A-10	Sequence 10, Appl
c	21	57	19.3	5649	3	US-09-902-540-735	Sequence 735, App
c	22	56	19.0	2561	3	US-09-616-289-48	Sequence 48, Appl
c	23	56	19.0	2561	3	US-09-976-740-48	Sequence 48, Appl
c	24	56	19.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
c	25	55.8	18.9	495	3	US-09-902-540-2340	Sequence 2340, Ap
	26	55.4	18.8	1018	2	US-08-444-083-7	Sequence 7, Appli
	27	55.4	18.8	1018	2	US-08-286-304-7	Sequence 7, Appli
	28	55.4	18.8	1018	2	US-08-442-745-7	Sequence 7, Appli
	29	55.4	18.8	1018	2	US-08-443-129-7	Sequence 7, Appli
	30	55.4	18.8	1018	2	US-08-443-952-7	Sequence 7, Appli
	31	55.4	18.8	1018	2	US-08-443-130-7	Sequence 7, Appli
	32	55.4	18.8	1018	3	US-08-898-911-7	Sequence 7, Appli
	33	55.4	18.8	1018	6	PCT-US95-04467-7	Sequence 7, Appli
c	34	55.4	18.8	1116	3	US-09-252-991A-2102	Sequence 2102, Ap
	35	55.4	18.8	1200	3	US-09-252-991A-1861	Sequence 1861, Ap
	36	55.4	18.8	1224	3	US-09-252-991A-1793	Sequence 1793, Ap
c	37	55.4	18.8	1722	3	US-09-252-991A-2173	Sequence 2173, Ap
c	38	55.2	18.7	789	3	US-09-902-540-7266	Sequence 7266, Ap
c	39	55.2	18.7	849	3	US-09-902-540-5481	Sequence 5481, Ap
c	40	55.2	18.7	4324	3	US-09-902-540-684	Sequence 684, App
	41	55.2	18.7	29899	3	US-09-902-540-1265	Sequence 1265, Ap
c	42	54.8	18.6	897	3	US-09-434-288-6	Sequence 6, Appli
c	43	54.6	18.5	71989	3	US-09-443-501A-2	Sequence 2, Appli
c	44	54.6	18.5	71989	3	US-09-724-889A-2	Sequence 2, Appli
c	45	54.6	18.5	71989	3	US-09-724-876-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-902-540-3603/c
; Sequence 3603, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3603
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3603

Query Match 21.8%; Score 64.4; DB 3; Length 1338;
Best Local Similarity 51.8%; Pred. No. 0.0099;
Matches 146; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              ||||| | || | ||  || |||  | | |||| ||   || |||| | | | | | ||
Db       1269 GGCGGCCAGCTCTTCCCGGGCCTGGGGCAGGCCCAGGTTGCCGGCGCCACCCAGGTGGGT 1210

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
                | |    | ||| |  ||   | ||              |  |   |    | |||  | |
Db       1209 GGCCGTCAGGGCGCCGCGGTCTGGACGGAACGTTCCTTCCACCAGCTCGCGGCCCACCTG 1150

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              || |  ||||||       ||| | | |  |  | |  | || |  || ||| ||||| |
Db       1149 CTTGTAGGCGTCACGGAAGGCCTGTCCCTGGGCCACCAGCGCATAGGCATGGTGCGCCGC 1090

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              | |||||| ||| ||| ||| ||| || |||| |  |  |||    |     |||  |
Db       1089 GTACAGCGTGTCGTCGCTGGCACGCGCCGCCGCGTCCGCCTTCACCTGGAGCGCGGGCAC 1030

Qy        246 CACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAG 287
              || | | |   || | || |   ||  | | |  | |||
Db       1029 CAACCGGGTGAGCACGTCCAGCAGCTCCCGCGTGGAGCCCAG 988
```

```
RESULT 2
US-09-902-540-1173/c
; Sequence 1173, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1173
;   LENGTH: 20113
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-1173

  Query Match             21.8%;  Score 64.4;  DB 3;  Length 20113;
  Best Local Similarity   51.8%;  Pred. No. 0.0067;
  Matches  146;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              ||||| | || | ||  || |||  | | |||| ||   || |||| | | | | | ||
Db      19270 GGCGGCCAGCTCTTCCCGGGCCTGGGGCAGGCCCAGGTTGCCGGCGCCACCCAGGTGGGT 19211

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
                | |    | ||| |  ||   | ||           |  |   |    | |||  | |
Db      19210 GGCCGTCAGGGCGCCGCGGTCTGGACGGAACGTTCCTTCCACCAGCTCGCGGCCCACCTG 19151

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              || |  ||||||       ||| | | |  | | |  | || |  || ||| ||||| |
Db      19150 CTTGTAGGCGTCACGGAAGGCCTGTCCCTGGGCCACCAGCGCATAGGCATGGTGCGCCGC 19091

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              | |||||| ||| |||  |||  ||| || |||| |  |  |||    |     |||  |
Db      19090 GTACAGCGTGTCGTCGCTGGCACGCGCCGCCGCGTCCGCCTTCACCTGGAGCGCGGGCAC 19031

Qy        246 CACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAG 287
              || | | |    || | || |    ||  | | |  | |||
Db      19030 CAACCGGGTGAGCACGTCCAGCAGCTCCCGCGTGGAGCCCAG 18989


RESULT 3
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             21.2%;  Score 62.4;  DB 3;  Length 71989;
  Best Local Similarity   51.4%;  Pred. No. 0.012;
  Matches  144;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         14 GCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGT 73
              ||||||| || |  |    |||| |||| || |  ||||      || || ||  |||
Db      65289 GCGCGTCGCGGCCGGCGTCGCGGCGGAGCTCGTGCACGGCCTCGGCCTGCGCCCGCGCGG 65348

Qy         74 CGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGG 133
                | ||| ||   | ||   || | |||||    ||  |  ||          | |
Db      65349 GATCACGGACGCCGACGCCGCCGCGCTGCGCGCGCTCCGCCCCCCGGACGGCGAGGGGCA 65408

Qy        134 CGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCG 193
              || ||   |   || ||||| ||||| ||    || ||||| || | ||  |||
Db      65409 CGGCGCCTTCGTCCGGGCGCTCGCGCGCGATCTCGCGCGCGCGGGGGGCGCCGGCGTCGC 65468

Qy        194 CGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCG 253
              |||| |||| | ||| | || || |  || |          ||   | |   | ||
Db      65469 CGTCGTCGGCGACGGCCAGCCGCCCATCGTCCACGCCCTCGGGCACGTCATCAACGCCGC 65528

Qy        254 CTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
                |||    |||| | | |  |  || ||| |  |||| |
Db      65529 GCTCCGCAGCCGGGCGGCCTGGATGGTCGATCCTGTGCTG 65568


RESULT 4
US-09-724-889A-2
; Sequence 2, Application US/09724889A
; Patent No. 6858411
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/724,889A
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: US/09/443,501A
;  PRIOR FILING DATE: 1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-724-889A-2

  Query Match             21.2%;  Score 62.4;  DB 3;  Length 71989;
  Best Local Similarity   51.4%;  Pred. No. 0.012;
  Matches  144;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;
```

```
Qy         14 GCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGT 73
              |||||||| || |  |    |||| |||| || |  ||||       || || ||  |||
Db      65289 GCGCGTCGCGGCCGGCGTCGCGGCGGAGCTCGTGCACGGCCTCGGCCTGCGCCCGCGCGG 65348

Qy         74 CGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGG 133
                | ||| ||   | ||   || | |||||    ||  |  ||          ||
Db      65349 GATCACGGACGCCGACGCCGCCGCGCTGCGCGCGCTCCGCCCCCCCGGACGGCGAGGGGCA 65408

Qy        134 CGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCG 193
              || ||   |   || ||||| ||||| ||    || ||||| || | ||  |||
Db      65409 CGGCGCCTTCGTCCGGGCGCTCGCGCGCGATCTCGCGCGCGCGGGGGGCGCCGGCGTCGC 65468

Qy        194 CGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCG 253
              |||| |||| | ||| | || || | || |         ||   | |    | ||
Db      65469 CGTCGTCGGCGACGGCCAGCCGCCCATCGTCCACGCCCTCGGGCACGTCATCAACGCCGC 65528

Qy        254 CTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
                |||    |||| | | |  |  || ||| |  |||| |
Db      65529 GCTCCGCAGCCGGGCGGCCTGGATGGTCGATCCTGTGCTG 65568
```

RESULT 5
US-09-724-876-2
; Sequence 2, Application US/09724876
; Patent No. 6921650
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/724,876
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/443,501
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-724-876-2

Query Match 21.2%; Score 62.4; DB 3; Length 71989;
Best Local Similarity 51.4%; Pred. No. 0.012;
Matches 144; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
Qy         14 GCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGT 73
              |||||||| || |  |    |||| |||| || |  ||||       || || ||  |||
Db      65289 GCGCGTCGCGGCCGGCGTCGCGGCGGAGCTCGTGCACGGCCTCGGCCTGCGCCCGCGCGG 65348

Qy         74 CGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGG 133
                | ||| ||   | ||   || | |||||    ||  |  ||          ||
Db      65349 GATCACGGACGCCGACGCCGCCGCGCTGCGCGCGCTCCGCCCCCCCGGACGGCGAGGGGCA 65408

Qy        134 CGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCG 193
              || ||   |   || ||||| ||||| ||    || ||||| || | ||  |||
Db      65409 CGGCGCCTTCGTCCGGGCGCTCGCGCGCGATCTCGCGCGCGCGGGGGGCGCCGGCGTCGC 65468

Qy        194 CGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCG 253
              |||| |||| | ||| | || || | || |         ||   | |    | ||
Db      65469 CGTCGTCGGCGACGGCCAGCCGCCCATCGTCCACGCCCTCGGGCACGTCATCAACGCCGC 65528

Qy        254 CTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
                |||    |||| | | |  |  || ||| |  |||| |
Db      65529 GCTCCGCAGCCGGGCGGCCTGGATGGTCGATCCTGTGCTG 65568
```

RESULT 6
US-09-758-759-148/c
; Sequence 148, Application US/09758759
; Patent No. 6861513
; GENERAL INFORMATION:
; APPLICANT: Hosted, Thomas J.
; APPLICANT: Wang, Tim X.
; APPLICANT: Horan, Ann C.
; TITLE OF INVENTION: Everninomicin Biosynthetic Genes
; FILE REFERENCE: ID0983K US
; CURRENT APPLICATION NUMBER: US/09/758,759
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/175,751
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 148
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1683)
; OTHER INFORMATION: evcD
US-09-758-759-148

Query Match 20.7%; Score 61.2; DB 3; Length 1683;
Best Local Similarity 52.8%; Pred. No. 0.032;
Matches 132; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              ||  |||  || |   | | |  ||  ||   | | |  | |  || |   |||||| ||
Db        613 CGTCGGCCATCCGGCGGACCACCTCGACCTGGCCCCGGGCCCGGCCGAACGCGCGCAGCG 554

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
              | | |  | || || |   | || | | | || ||  |||||  |    |||| ||| |
Db        553 TGGGCAGGCCGGCGACCATGTCGAGGAAGTGCCCGCCGAGCGTCGACAGCCGCCGCCACT 494

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
              |||||||  |||| |  |  |  || ||  |  || | | |  | || |   ||||
Db        493 GCCGCTCGGTGGCGGCCTGCGCCTGCCAGCCGAGCAGCGCCCCGAAGACCGGGATCAGCG 434

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |  || |||||||  |   |  ||| |    |  |||||  ||| |  | | | |
Db        433 GCAGGGTCAGCGCGACGATGACGGCCGAGCCCCAGTCGGCGAAGGTGATCCGGGCCAGCA 374

Qy        242 GGCCCACCTG 251
                | |||||  |
Db        373 CGGCCACCGG 364
```

RESULT 7
US-09-477-962-1/c
; Sequence 1, Application US/09477962
; Patent No. 6927286
; GENERAL INFORMATION:
; APPLICANT: SHEN, BEN
; APPLICANT: DU, LIANGCHENG
; APPLICANT: SANCHEZ, CESAR
; APPLICANT: CHEN, MEI
; APPLICANT: EDWARDS, DANIEL J.
; TITLE OF INVENTION: BLEOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES
; FILE REFERENCE: 407T-895820US
; CURRENT APPLICATION NUMBER: US/09/477,962

```
;  CURRENT FILING DATE:  2000-01-05
;  PRIOR APPLICATION NUMBER: 60/115,435
;  PRIOR FILING DATE: 1999-01-06
;  PRIOR APPLICATION NUMBER: 60/118,848
;  PRIOR FILING DATE: 1999-02-05
;  NUMBER OF SEQ ID NOS: 133
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1
;   LENGTH: 58857
;   TYPE: DNA
;   ORGANISM: Streptomyces verticillus
US-09-477-962-1

  Query Match             20.7%;  Score 61.2;  DB 3;  Length 58857;
  Best Local Similarity   52.3%;  Pred. No. 0.019;
  Matches  135;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          31 GTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGCAGG 90
               || | |||| |   |||||||||| | |       |||| ||  |  |  |  ||    |
Db       42320 GTTCCGCCGCGACTGCGGGCGGCGAGGAGGGAGTGCCCGCCGAGGAGGAAGAAGTCGTCG 42261

Qy          91 CGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGG 150
                |||||  || ||   |  || |      |  ||   ||| | |||||     ||  ||
Db       42260 TGCGCGCCGACCGGCCCCTCCAGGCCGAGCACCCCGGCGAAGACGTCGGCGACCGTACGC 42201

Qy         151 CGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGC 210
                | |   |||||  |   |   ||   |           |  |   ||||||||||||
Db       42200 TCGGTGGGGGTGCGGGGGGCGGTCGGGGCGGGGCGCCCGGCCCGGGGGTCGGGGGCGGGC 42141

Qy         211 GGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCG 270
               || |||||  | |||||  |||||     ||   ||  | || | |||||||| |   ||
Db       42140 AGCCGCGCGTGGTCGGTCTTGCCGTCGCGGTCACCGGCAGCTCGTCCAGGACGGTGACG 42081

Qy         271 TCGTGGGGGCCGAGCAGG 288
               ||  |||  ||||  ||
Db       42080 GCGGTGGGCACGAGGTGG 42063


RESULT 8
US-09-758-759-1/c
; Sequence 1, Application US/09758759
; Patent No. 6861513
; GENERAL INFORMATION:
;  APPLICANT: Hosted, Thomas J.
;  APPLICANT:  Wang, Tim X.
;  APPLICANT:  Horan, Ann C.
;  TITLE OF INVENTION: Everninomicin Biosynthetic Genes
;  FILE REFERENCE: ID0983K US
;  CURRENT APPLICATION NUMBER: US/09/758,759
;  CURRENT FILING DATE:  2001-01-11
;  PRIOR APPLICATION NUMBER: US 60/175,751
;  PRIOR FILING DATE: 2000-01-12
;  NUMBER OF SEQ ID NOS: 204
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 109519
;   TYPE: DNA
;   ORGANISM: Micromonospora carbonacea
US-09-758-759-1

  Query Match             20.7%;  Score 61.2;  DB 3;  Length 109519;
  Best Local Similarity   52.8%;  Pred. No. 0.018;
  Matches  132;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy           2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
               ||  |||  || |   | | |  ||  ||    |  | |   |  || |   ||||| ||
Db       92760 CGTCGGCCATCCGGCGGACCACCTCGACCTGGCCCCGGGCCCGGCCGAACGCGCGCAGCG 92701

Qy          62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
               | | |  | || || |   | || | | | || ||  |||||  |    |||| ||| |
Db       92700 TGGGCAGGCCGGCGACCATGTCGAGGAAGTGCCCGCCGAGCGTCGACAGCCGCCGCCACT 92641

Qy         122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
               |||||||  |||| |   |  |  || ||  |   || | | |  | || |   ||||
Db       92640 GCCGCTCGGTGGCGGCCTGCGCCTGCCAGCCGAGCAGCGCCCCGAAGACCGGGATCAGCG 92581

Qy         182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
                 | || ||||||| |   |  ||| |    |    |||||  ||| |  | | | |
Db       92580 GCAGGGTCAGCGCGACGATGACGGCCGAGCCCCAGTCGGCGAAGGTGATCCGGGCCAGCA 92521

Qy         242 GGCCCACCTG 251
               | |||||  |
Db       92520 CGGCCACCGG 92511


RESULT 9
US-08-690-473-1/c
; Sequence 1, Application US/08690473
; Patent No. 5876923
;  GENERAL INFORMATION:
;    APPLICANT:  Leopardi, Rosario
;    APPLICANT:  Roizman, Bernard
;    TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 AS AN
;    TITLE OF INVENTION:  INHIBITOR OF APOPTOSIS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/690,473
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARCD:239
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-690-473-1

  Query Match             20.6%;  Score 60.8;  DB 2;  Length 4257;
  Best Local Similarity   53.6%;  Pred. No. 0.033;
  Matches  148;  Conservative    0;  Mismatches  127;  Indels    1;  Gaps    1;

Qy          13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
               | |||| ||  | |  |||     ||   ||  ||| | | | ||| ||  ||    | |
Db        2426 TCCGCGGCGGCGGCGGCCGCGGAGCTCGGCAGGCGCGGGTCCCGCGGCAGCGCGGGGCCC 2367

Qy          73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
               | | | ||||  ||||| | ||  | || |||||||||    |   ||   |   ||  |
Db        2366 AGGGCCCCGGCGACCAGGCTCACGGCGCGCACGGCGGCCACGGCGGCCTCGCTGCCGCCG 2307

Qy         133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
               ||  ||  | |  ||| ||||| |||| | |  |  ||||||  | || |       ||||
Db        2306 GCCACGCGCAGGTCCCCGCGCAGGCGCATGAGCACCAGCGCGTCGCGCACGAACCGCAGC 2247

Qy         193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
               ||  |      ||| || |  | |  ||| ||| |  || | |  | ||  | | |
Db        2246 TCGCGCAGCCACGCGCGCAGGCGGGGCGCGTCGG-CGTGCGGCGGCGGCGGGGAAGCGGG 2188
```

```
Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGG 288
              ||   | || ||   | | || |||||  |  || ||
Db       2187 GCCCGCGGGTCCCTCCGGCCGCGGGGGGCTGGCGGG 2152


RESULT 10
US-09-259-821A-1/c
; Sequence 1, Application US/09259821A
; Patent No. 6210926
; GENERAL INFORMATION:
;  APPLICANT:  LEOPARDI, ROSARIO
;  APPLICANT:  ROIZMAN, BERNARD
;  TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
;  FILE REFERENCE:  ARCD:317
;  CURRENT APPLICATION NUMBER: US/09/259,821A
;  CURRENT FILING DATE:  1999-03-01
;  PRIOR APPLICATION NUMBER:  08/690,473
;  PRIOR FILING DATE:  1996-07-26
;  NUMBER OF SEQ ID NOS:  2
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4257
;   TYPE: DNA
;   ORGANISM: HERPES VIRUS, TYPE 1
US-09-259-821A-1

  Query Match             20.6%;  Score 60.8;  DB 3;  Length 4257;
  Best Local Similarity   53.6%;  Pred. No. 0.033;
  Matches  148;  Conservative    0;  Mismatches  127;  Indels    1;  Gaps    1;

Qy         13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
              | |||| ||  | |  |||    ||   ||  ||| | | | ||| ||  ||    | |
Db       2426 TCCGCGGCGGCGGCGGCCGCGGAGCTCGGCAGGCGCGGGTCCCGCGGCAGCGCGGGGCCC 2367

Qy         73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
                | | | ||||  ||||| | ||  | || ||||||||    |   ||   |   || |
Db       2366 AGGGCCCCGGCGACCAGGCTCACGGCGCGCACGGCGGCCACGGCGGCCTCGCTGCCGCCG 2307

Qy        133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
              ||  ||  | |  ||| ||||| ||||  | |  |  ||||| | || |      ||||
Db       2306 GCCACGCGCAGGTCCCCGCGCAGGCGCATGAGCACCAGCGCGTCGCGCACGAACCGCAGC 2247

Qy        193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
              ||  |      ||| || |  | |  ||| ||| |  || |  |  || ||   | ||
Db       2246 TCGCGCAGCCACGCGCGCAGGCGGGGCGCGTCGG-CGTGCGGCGGCGGCGGGGAAGCGGG 2188

Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGG 288
              ||   | || ||   | | || |||||  |  || ||
Db       2187 GCCCGCGGGTCCCTCCGGCCGCGGGGGGCTGGCGGG 2152


RESULT 11
US-08-843-659-1/c
; Sequence 1, Application US/08843659
; Patent No. 6218103
;  GENERAL INFORMATION:
;    APPLICANT:  Leopardi, Roasrio
;    APPLICANT:  Roizman, Bernard
;    TITLE OF INVENTION:  HERPES SIMPLEX VIRUS US3 AND ICP4 AS
;    TITLE OF INVENTION:  INHIBITORS OF APOPTOSIS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/843,659
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARSB:519
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-843-659-1

  Query Match             20.6%;  Score 60.8;  DB 3;  Length 4257;
  Best Local Similarity   53.6%;  Pred. No. 0.033;
  Matches  148;  Conservative    0;  Mismatches  127;  Indels    1;  Gaps    1;

Qy         13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
              | |||| ||  | |  |||    ||   ||  ||| | | | ||| ||  ||    | |
Db       2426 TCCGCGGCGGCGGCGGCCGCGGAGCTCGGCAGGCGCGGGTCCCGCGGCAGCGCGGGGCCC 2367

Qy         73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
                | | | ||||  ||||| | ||  | || ||||||||    |   ||   |   || |
Db       2366 AGGGCCCCGGCGACCAGGCTCACGGCGCGCACGGCGGCCACGGCGGCCTCGCTGCCGCCG 2307

Qy        133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
              ||  ||  | |  ||| ||||| ||||  | |  |  ||||| | || |      ||||
Db       2306 GCCACGCGCAGGTCCCCGCGCAGGCGCATGAGCACCAGCGCGTCGCGCACGAACCGCAGC 2247

Qy        193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
              ||  |      ||| || |  | |  ||| ||| |  || |  |  || ||   | ||
Db       2246 TCGCGCAGCCACGCGCGCAGGCGGGGCGCGTCGG-CGTGCGGCGGCGGCGGGGAAGCGGG 2188

Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGG 288
              ||   | || ||   | | || |||||  |  || ||
Db       2187 GCCCGCGGGTCCCTCCGGCCGCGGGGGGCTGGCGGG 2152


RESULT 12
US-09-825-288A-1/c
; Sequence 1, Application US/09825288A
; Patent No. 6723511
; GENERAL INFORMATION:
;  APPLICANT: LEOPARDI, ROSARIO
;  APPLICANT:  ROIZMAN, BERNARD
;  TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
;  FILE REFERENCE: ARCD:317USC1
;  CURRENT APPLICATION NUMBER: US/09/825,288A
;  CURRENT FILING DATE:  2001-04-02
;  PRIOR APPLICATION NUMBER: 09/259,821
;  PRIOR FILING DATE: 1999-03-01
;  PRIOR APPLICATION NUMBER: 08/690,473
;  PRIOR FILING DATE: 1996-07-26
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4257
;   TYPE: DNA
;   ORGANISM: HERPES VIRUS, TYPE 1
US-09-825-288A-1

  Query Match             20.6%;  Score 60.8;  DB 3;  Length 4257;
  Best Local Similarity   53.6%;  Pred. No. 0.033;
  Matches  148;  Conservative    0;  Mismatches  127;  Indels    1;  Gaps    1;

Qy         13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
              | |||| ||  | |  |||    ||   ||  ||| | | | ||| ||  ||    | |
```

```
Db       2426 TCCGCGGCGGCGGCGGCCGCGGAGCTCGGCAGGCGCGGGTCCCGCGGCAGCGCGGGGCCC 2367

Qy         73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
              | | | ||||  ||||| | ||  | || ||||||||    |    ||   |    ||  |
Db       2366 AGGGCCCCGGCGACCAGGCTCACGGCGCGCACGGCGGCCACGGCGGCCTCGCTGCCGCCG 2307

Qy        133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
              ||  ||  | |  ||| ||||| ||||  | |  |  ||||||| | || |      ||||
Db       2306 GCCACGCGCAGGTCCCCGCGCAGGCGCATGAGCACCAGCGCGTCGCGCACGAACCGCAGC 2247

Qy        193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
              ||  |      ||| || |  | |  ||| ||| |  || |  |  | ||   | | |
Db       2246 TCGCGCAGCCACGCGCGCAGGCGGGGCGCGTCGG-CGTGCGGCGGCGGCGGGGAAGCGGG 2188

Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGG 288
              ||   | || ||   | | || |||||  |  || ||
Db       2187 GCCCGCGGGTCCCTCCGGCCGCGGGGGGCTGGCGGG 2152


RESULT 13
US-08-458-568A-11
; Sequence 11, Application US/08458568A
; Patent No. 5821339
;  GENERAL INFORMATION:
;    APPLICANT:  Schaffer, Priscilla A.
;    APPLICANT:  Yeh, Lily
;    TITLE OF INVENTION:  Compositions and Methods for Treatment of Herpesvirus
;    TITLE OF INVENTION:  Infections
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,568A
;      FILING DATE:  02-JUNE-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/065,146
;      FILING DATE:  05-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leary Ph.D., Kathryn R.
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  DFCI-0029
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpes simplex virus
;      STRAIN:  Herpes Simplex Virus Type 1
US-08-458-568A-11

  Query Match           20.6%;  Score 60.8;  DB 2;  Length 12001;
  Best Local Similarity   53.6%;  Pred. No. 0.028;
  Matches  148;  Conservative    0;  Mismatches  127;  Indels    1;  Gaps    1;

Qy         13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
              | |||| ||  | |  |||      ||   ||  ||| | | | ||| ||  ||    | |
Db       4248 TCCGCGGCGGCGGCGGCCGCGGAGCTCGGCAGGCGCGGGTCCCGCGGCAGCGCGGGGCCC 4307

Qy         73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
              | | | ||||  ||||| | ||  | || ||||||||    |    ||   |    ||  |
Db       4308 AGGGCCCCGGCGACCAGGCTCACGGCGCGCACGGCGGCCACGGCGGCCTCGCTGCCGCCG 4367

Qy        133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
              ||  ||  | |  ||| ||||| ||||  | |  |  ||||||| | || |      ||||
Db       4368 GCCACGCGCAGGTCCCCGCGCAGGCGCATGAGCACCAGCGCGTCGCGCACGAACCGCAGC 4427

Qy        193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
              ||  |      ||| || |  | |  ||| ||| |  || |  |  | ||   | | |
Db       4428 TCGCGCAGCCACGCGCGCAGGCGGGGCGCGTCGG-CGTGCGGCGGCGGCGGGGAAGCGGG 4486

Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGG 288
              ||   | || ||   | | || |||||  |  || ||
Db       4487 GCCCGCGGGTCCCTCCGGCCGCGGGGGGCTGGCGGG 4522


RESULT 14
US-08-726-306A-52
; Sequence 52, Application US/08726306A
; Patent No. 5958684
;  GENERAL INFORMATION:
;    APPLICANT:  van Leeuwen, Frederik Willem
;    APPLICANT:  Burbach, Johannes Peter Henri
;    APPLICANT:  Grosveld, Franklin G.
;    TITLE OF INVENTION:  DIAGNOSIS METHOD AND REAGENTS
;    NUMBER OF SEQUENCES:  189
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  1 Financial Center
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,306A
;      FILING DATE:  02-Oct-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 95/20080.4
;      FILING DATE:  02-Oct-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/009,832
;      FILING DATE:  01-Jan-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Ph.D., Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  96,048-A (3255/00784)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 345-9100
;      TELEFAX:  (617) 345-9111
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1162 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-726-306A-52

  Query Match           20.3%;  Score 59.8;  DB 2;  Length 1162;
  Best Local Similarity   54.4%;  Pred. No. 0.057;
  Matches  141;  Conservative    0;  Mismatches  117;  Indels    1;  Gaps    1;
```

```
Qy          36 GCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGC-AGGCGCG 94
               || | |   || |  |||| |||||| ||    ||    || |   | | || |||  ||
Db         672 GCTGCGGCAGCAGCAGGCGGGCCGCTCCGCTATGGGCGAGCTGTACGAGCGCGAGGTCCG 731

Qy          95 CGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCA 154
               ||||  ||||||||    || || ||         | | |||   | || |||| || |||
Db         732 CGAGATGCGCGGCGCGGTGCTGCGCCTGGGCGCGGCGCGCGGTCAGCTACGCCTGGAGCA 791

Qy         155 CGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCG 214
               | || |  |  || |  | | | || | |  | |||||| |   ||  |  |  ||||
Db         792 GGAGCACCTGCTCGAGGACATCGCGCACGTGCGCCAGCGCCTAGACGACGAGGCCCGGCA 851

Qy         215 GCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGT 274
               ||| |  |  |  |||| |  |   | || |   |||||||  | |   | |||  |
Db         852 GCGAGAGGAGGCCGAGGCGGCGGCCCGCGCGCTGGCGCGCTTCGCGCAGGAGGCCGAGGC 911

Qy         275 GGGGGCCGAGCAGGTGCAG 293
               || |  || | |  |||||
Db         912 GGCGCGCGTGGACCTGCAG 930


RESULT 15
US-10-237-551-193/c
; Sequence 193, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
;  APPLICANT: Day, Craig H.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Parsons, Joseph M.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.538C3
;  CURRENT APPLICATION NUMBER: US/10/237,551
;  CURRENT FILING DATE:  2002-09-06
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 193
;   LENGTH: 3957
;   TYPE: DNA
;   ORGANISM: HSV2
US-10-237-551-193

  Query Match            20.2%;  Score 59.6;  DB 3;  Length 3957;
  Best Local Similarity  50.7%;  Pred. No. 0.052;
  Matches  143;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy          13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
               | |||| ||  | |  | |    ||  |||  ||| | |   ||| ||  || |  | |
Db        2273 TCCGCGGCGGCGGCGGCGGCGGAGCTCAGCAGGCGCGGGCTCCGCGGCAGCGCCGGGCCC 2214

Qy          73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
                 | | | ||||  ||||| | ||  | || ||||||||   |   ||   |   ||  |
Db        2213 AGGGCCCCGGCGACCAGGCTCACGGCGCGCACGGCGGCCACGGCGGCCTCGCTGCCGCCG 2154

Qy         133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
               ||  ||  | |  ||| ||||| ||||    |  |  |||||| | || |       ||||
Db        2153 GCCACGCGCAGGTCCCCGCGCAGGCGCATCAGCACCAGCGCGTCGCGCACGAACCGCAGC 2094

Qy         193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
                ||  |     |||| || |  | |  ||| |||   |       |  ||   |  ||
Db        2093 TCGCGCAGCCAGGCGCGCAGGCGGGGCGCGTCGGCGTGCGGCGGGGCGGCCGCGCCCGCG 2034

Qy         253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGC 294
               |   || ||  |||     ||  ||| | |     |  ||||
Db        2033 GGCCCCGGGCGCGGGGGCGCGGCGGGCCGGGCTCCGGCCAGC 1992


Search completed: April 27, 2006, 12:42:42
Job time : 61.863 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:42:50 ; Search time 185.244 Seconds
(without alignments)
6467.537 Million cell updates/sec

Title: US-10-729-802-1_COPY_77000_77294
Perfect score: 295
Sequence: 1 gcgagggcggtctgcgcgtc.....ggggccgagcaggtgcagcg 295

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	63.8	21.6	1047	7	US-10-496-351-11	Sequence 11, Appl
2	63.8	21.6	82746	7	US-10-496-351-56	Sequence 56, Appl
c 3	60.4	20.5	9975	7	US-10-496-351-4	Sequence 4, Appli
4	59.8	20.3	3695	14	US-11-169-041-84	Sequence 84, Appl
5	59.8	20.3	121736	11	US-11-114-798-49	Sequence 49, Appl
6	59.8	20.3	318488	11	US-11-114-798-58	Sequence 58, Appl
c 7	57.6	19.5	2080	14	US-11-140-625-2	Sequence 2, Appli
c 8	57.2	19.4	1500	9	US-10-297-056-10	Sequence 10, Appl
9	56.8	19.3	153376	14	US-11-121-086-5	Sequence 5, Appli
10	56.8	19.3	172543	14	US-11-121-086-6	Sequence 6, Appli
11	56.4	19.1	9975	7	US-10-496-351-4	Sequence 4, Appli
c 12	56.4	19.1	82746	7	US-10-496-351-56	Sequence 56, Appl
13	55.4	18.8	663	10	US-10-301-480-566618	Sequence 566618,
14	55.4	18.8	663	10	US-10-301-480-1180027	Sequence 1180027,
c 15	55.4	18.8	14055	7	US-10-496-351-1	Sequence 1, Appli
c 16	55.2	18.7	37507	8	US-10-522-037-2	Sequence 2, Appli
c 17	54	18.3	687	9	US-10-714-887-89	Sequence 89, Appl
c 18	53.8	18.2	4770	7	US-10-496-351-3	Sequence 3, Appli
c 19	53.6	18.2	113193	7	US-10-478-943D-1	Sequence 1, Appli
20	53	18.0	14055	7	US-10-496-351-1	Sequence 1, Appli
21	52.6	17.8	1247	14	US-11-226-555-32	Sequence 32, Appl
c 22	52.4	17.8	975	11	US-11-096-568A-9234	Sequence 9234, Ap
c 23	52.4	17.8	1047	7	US-10-496-351-11	Sequence 11, Appl
c 24	52.2	17.7	3354	7	US-10-496-351-12	Sequence 12, Appl
c 25	51.8	17.6	9222	7	US-10-496-351-2	Sequence 2, Appli
c 26	51.6	17.5	610	6	US-09-925-065A-869021	Sequence 869021,
c 27	51.6	17.5	1694969	7	US-10-506-454-1690	Sequence 1690, Ap
c 28	51.4	17.4	88421	14	US-11-205-109-1	Sequence 1, Appli
c 29	51.4	17.4	295772	9	US-10-330-773-410	Sequence 410, App
c 30	51.2	17.4	381	6	US-09-925-065A-545041	Sequence 545041,
c 31	51.2	17.4	610	6	US-09-925-065A-869022	Sequence 869022,
c 32	51.2	17.4	1734	11	US-11-096-568A-12766	Sequence 12766, A
c 33	51	17.3	1491	11	US-11-096-568A-27218	Sequence 27218, A
c 34	50.8	17.2	381	6	US-09-925-065A-545042	Sequence 545042,
c 35	50.8	17.2	381	6	US-09-925-065A-545043	Sequence 545043,
36	50.8	17.2	687411	9	US-10-330-773-26	Sequence 26, Appl
37	50.6	17.2	46752	8	US-10-995-561-13410	Sequence 13410, A
38	50.6	17.2	415117	8	US-10-995-561-13274	Sequence 13274, A
c 39	50.4	17.1	1140	11	US-11-096-568A-16700	Sequence 16700, A
c 40	50.4	17.1	1306	11	US-11-096-568A-24130	Sequence 24130, A
41	50.4	17.1	195998	8	US-10-995-561-13489	Sequence 13489, A
c 42	50.2	17.0	974	10	US-10-301-480-599183	Sequence 599183,
c 43	50.2	17.0	974	10	US-10-301-480-1212592	Sequence 1212592,
c 44	50	16.9	702	9	US-10-714-887-99	Sequence 99, Appl
c 45	50	16.9	3018	7	US-10-496-351-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-10-496-351-11
; Sequence 11, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-11

Query Match 21.6%; Score 63.8; DB 7; Length 1047;
Best Local Similarity 51.6%; Pred. No. 0.00022;
Matches 146; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
Qy          13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
               | | || || ||| | ||| | || || || | || | || ||| |
Db         458 TCCACGACGCGGTGGCCACCGGCCCCGGCGACCCGTTCGACGAGGCGATCCGCTCCGCGC 517

Qy          73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
               ||| |  | || |  |||| | ||  || |||  ||| |  ||| |
Db         518 TCGACCTGACCGGGGCCGCGCGGGTGGTCGACGTCGGCGGCGGGCGGGGCAGGCTGCTCG 577

Qy         133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
                ||    |||||  |   | || |||| |||||||||  | || |  | ||   |
Db         578 CCGCGCTGCTGCGCGAGCACCCCGTGCTCTCGGGCGTGCTGTTCGACCTGCCCGAGGCGG 637

Qy         193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
               |     || ||       || |||||||| ||  | || |       ||||   |  |
Db         638 TCCGGGACGCGGTTCCCGAGCTGCGCGGCGACGACCGGGTCCAGATCAGGGCGGGCGACC 697
```

```
Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              ||| |||||||  || || || || ||   | |||| |||
Db        698 TGTTCGCGGACCGGGTCGGCGCGGCGGACGTCTACGTGCTGCG 740


RESULT 2
US-10-496-351-56
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
;   LENGTH: 82746
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

  Query Match             21.6%;  Score 63.8;  DB 7;  Length 82746;
  Best Local Similarity   51.6%;  Pred. No. 4.8e-05;
  Matches  146;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy         13 TGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCG 72
              | | || || |||   |      |||    | ||  || || | || | ||  |||
Db      76046 TCCACGACGCGGTGGCCACCGGCCCCGGCGACCCGTTCGACGAGGCGATCCGCTCCGCGC 76105

Qy         73 TCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGG 132
              ||| |  |  || |   ||||| | ||    || ||||    |||        |||    |
Db      76106 TCGACCTGACCGGGGCCGCGCGGGTGGTCGACGTCGGCGGCGGGCGGGGCAGGCTGCTCG 76165

Qy        133 GCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGC 192
               ||     ||||||  |   | || |||| |||||||||   | || |  | ||    |
Db      76166 CCGCGCTGCTGCGCGAGCACCCCGTGCTCTCGGGCGTGCTGTTCGACCTGCCCGAGGCGG 76225

Qy        193 GCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGC 252
              |    || ||       || |||||||| ||  | || |      ||||   |   |
Db      76226 TCCGGGACGCGGTTCCCGAGCTGCGCGGCGACGACCGGGTCCAGATCAGGGCGGGCGACC 76285

Qy        253 GCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |||  |||||||  || || || || ||   | |||| |||
Db      76286 TGTTCGCGGACCGGGTCGGCGCGGCGGACGTCTACGTGCTGCG 76328


RESULT 3
US-10-496-351-4/c
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 9975
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-4

  Query Match             20.5%;  Score 60.4;  DB 7;  Length 9975;
  Best Local Similarity   50.7%;  Pred. No. 0.00046;
  Matches  145;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              ||| ||||  |  ||||  |  |    |    ||   |||| ||| |    ||      |
Db       7150 CGACGGCGACCATCGCGCCCGGTCGCAGGTCGCGCATGAGCACGCCGCGCTCGGCGACCA 7091

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
              ||  |  ||  || || | | |  |  || ||||| | ||||||||||    |   ||
Db       7090 GCGCGCACGCCGCGTCGAGCGGGAACGCGCCCGCGACGTGCGCGGCGGCGATCTCGCCGA 7031

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
              |   | |   || |    | || | | |  ||  | |||     || |||   ||||
Db       7030 CCGAGTGGCCCGCCACCGCGTCCGGCGCGAGGCCGAACGCGCCCAGCTGGGCGGCCAGCG 6971

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |     ||| |||  | ||| ||  |||   ||  || ||||  |||||||  |
Db       6970 CGGTCTGCAGGGCGAACGCGGCGGTCTGCGCGTGCAGGGTGCCGTCCAGGCCGGGGCGCT 6911

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAG 287
              ||||  |    |  |  | |||||    |||  | |||||   ||
Db       6910 CCCCCAGCGCGACGGCGCGCACCGGGTGCTCGGCGCGGCCGTCGAG 6865


RESULT 4
US-11-169-041-84
; Sequence 84, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
;  TITLE OF INVENTION:  COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
;  TITLE OF INVENTION:  KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 10001 NP
;  CURRENT APPLICATION NUMBER: US/11/169,041
;  CURRENT FILING DATE:  2005-06-28
;  PRIOR APPLICATION NUMBER: 60/584,405
;  PRIOR FILING DATE: 2004-06-30
;  NUMBER OF SEQ ID NOS: 527
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 84
;   LENGTH: 3695
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-169-041-84

  Query Match             20.3%;  Score 59.8;  DB 14;  Length 3695;
  Best Local Similarity   54.4%;  Pred. No. 0.00086;
  Matches  141;  Conservative    0;  Mismatches  117;  Indels    1;  Gaps    1;

Qy         36 GCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGC-AGGCGCG 94
              || | |   || |  |||| |||||| ||    ||    || |   | | || |||  ||
Db       426 GCTGCGGCAGCAGCAGGCGGGCCGCTCCGCTATGGGCGAGCTGTACGAGCGCGAGGTCCG 485

Qy         95 CGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCA 154
              ||||  ||||||||     || || ||         | | |||    | || |||| || |||
Db       486 CGAGATGCGCGGCGCGGTGCTGCGCCTGGGCGCGGCGCGCGGTCAGCTACGCCTGGAGCA 545

Qy        155 CGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCG 214
              | || |  |  || |  | | | || | |  | |||||| |   ||  |  |  ||||
Db       546 GGAGCACCTGCTCGAGGACATCGCGCACGTGCGCCAGCGCCTAGACGACGAGGCCCGGCA 605

Qy        215 GCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGT 274
              ||| |  |  |   |||| |  |   | || |    |||||||  | |   | |||  |
Db       606 GCGAGAGGAGGCCGAGGCGGCGGCCCGCGCGCTGGCGCGCTTCGCGCAGGAGGCCGAGGC 665

Qy        275 GGGGGCCGAGCAGGTGCAG 293
              || |  || | |  |||||
Db       666 GGCGCGCGTGGACCTGCAG 684


RESULT 5
```

```
US-11-114-798-49
; Sequence 49, Application US/11114798
; Publication No. US20060035246A1
; GENERAL INFORMATION:
;  APPLICANT: WU, RINA
;  APPLICANT:  MARQUEZ, ABBEY
;  TITLE OF INVENTION: CHROMOGENIC IN SITU HYBRIDIZATION METHODS, KITS, AND
;  TITLE OF INVENTION:  COMPOSITIONS
;  FILE REFERENCE: 0618.011.0004
;  CURRENT APPLICATION NUMBER: US/11/114,798
;  CURRENT FILING DATE:  2005-04-26
;  PRIOR APPLICATION NUMBER: 10/173,525
;  PRIOR FILING DATE: 2002-06-17
;  PRIOR APPLICATION NUMBER: 09/952,851
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 49
;   LENGTH: 121736
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: nucleotide sequence
US-11-114-798-49

  Query Match            20.3%;  Score 59.8;  DB 11;  Length 121736;
  Best Local Similarity  54.4%;  Pred. No. 0.00025;
  Matches  141;  Conservative    0;  Mismatches  117;  Indels    1;  Gaps    1;

Qy         36 GCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGC-AGGCGCG 94
              || | |   || |  |||| |||||| ||    ||    || |   | | || |||  ||
Db      47118 GCTGCGGCAGCAGCAGGCGGGCCGCTCCGCTATGGGCGAGCTGTACGAGCGCGAGGTCCG 47177

Qy         95 CGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCA 154
              ||||  ||||||||    || || ||         | | |||   | || |||| || |||
Db      47178 CGAGATGCGCGGCGCGGTGCTGCGCCTGGGCGCGGCGCGCGGTCAGCTACGCCTGGAGCA 47237

Qy        155 CGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCG 214
               | || |  |  || |  | | | || | |  | ||||| |   ||  |  |  ||||
Db      47238 GGAGCACCTGCTCGAGGACATCGCGCACGTGCGCCAGCGCCTAGACGACGAGGCCCGGCA 47297

Qy        215 GCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGT 274
              ||| |  |  |   |||| |  |  | || |    ||||||  | |   | |||  |
Db      47298 GCGAGAGGAGGCCGAGGCGGCGGCCCGCGCGCTGGCGCGCTTCGCGCAGGAGGCCGAGGC 47357

Qy        275 GGGGGCCGAGCAGGTGCAG 293
              || |  || | |  |||||
Db      47358 GGCGCGCGTGGACCTGCAG 47376


RESULT 6
US-11-114-798-58
; Sequence 58, Application US/11114798
; Publication No. US20060035246A1
; GENERAL INFORMATION:
;  APPLICANT: WU, RINA
;  APPLICANT:  MARQUEZ, ABBEY
;  TITLE OF INVENTION: CHROMOGENIC IN SITU HYBRIDIZATION METHODS, KITS, AND
;  TITLE OF INVENTION:  COMPOSITIONS
;  FILE REFERENCE: 0618.011.0004
;  CURRENT APPLICATION NUMBER: US/11/114,798
;  CURRENT FILING DATE:  2005-04-26
;  PRIOR APPLICATION NUMBER: 10/173,525
;  PRIOR FILING DATE: 2002-06-17
;  PRIOR APPLICATION NUMBER: 09/952,851
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 58
;   LENGTH: 318488
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: nucleotide sequence
US-11-114-798-58

  Query Match            20.3%;  Score 59.8;  DB 11;  Length 318488;
  Best Local Similarity  54.4%;  Pred. No. 0.00018;
  Matches  141;  Conservative    0;  Mismatches  117;  Indels    1;  Gaps    1;

Qy         36 GCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGC-AGGCGCG 94
              || | |   || |  |||| |||||| ||    ||    || |   | | || |||  ||
Db       8566 GCTGCGGCAGCAGCAGGCGGGCCGCTCCGCTATGGGCGAGCTGTACGAGCGCGAGGTCCG 8625

Qy         95 CGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCA 154
              ||||  ||||||||    || || ||         | | |||   | || |||| || |||
Db       8626 CGAGATGCGCGGCGCGGTGCTGCGCCTGGGCGCGGCGCGCGGTCAGCTACGCCTGGAGCA 8685

Qy        155 CGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCG 214
               | || |  |  || |  | | | || | |  | ||||| |   ||  |  |  ||||
Db       8686 GGAGCACCTGCTCGAGGACATCGCGCACGTGCGCCAGCGCCTAGACGACGAGGCCCGGCA 8745

Qy        215 GCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGT 274
              ||| |  |  |   |||| |  |  | || |    ||||||  | |   | |||  |
Db       8746 GCGAGAGGAGGCCGAGGCGGCGGCCCGCGCGCTGGCGCGCTTCGCGCAGGAGGCCGAGGC 8805

Qy        275 GGGGGCCGAGCAGGTGCAG 293
              || |  || | |  |||||
Db       8806 GGCGCGCGTGGACCTGCAG 8824


RESULT 7
US-11-140-625-2/c
; Sequence 2, Application US/11140625
; Publication No. US20060026706A1
; GENERAL INFORMATION:
;  APPLICANT: Van Wezel, Gilles P.
;  APPLICANT:  Vijgenboom, Erik
;  TITLE OF INVENTION: A method for marker-less integration of a sequence of interest
;  TITLE OF INVENTION:  into the genome of a cell
;  FILE REFERENCE: P62377SUS00
;  CURRENT APPLICATION NUMBER: US/11/140,625
;  CURRENT FILING DATE:  2005-05-27
;  PRIOR APPLICATION NUMBER: EP 02080000.9
;  PRIOR FILING DATE: 2002-11-28
;  PRIOR APPLICATION NUMBER: PCT/NL2003/000835
;  PRIOR FILING DATE: 2003-11-27
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
;   LENGTH: 2080
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Sequence of the glkA region amplified from the S. coelicolor M145
;   OTHER INFORMATION: genome
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2080)
;   OTHER INFORMATION: Sequence starts at nucleotide position 37 and ends at nucleotide
;   OTHER INFORMATION: position 2116
US-11-140-625-2

  Query Match            19.5%;  Score 57.6;  DB 14;  Length 2080;
  Best Local Similarity  50.0%;  Pred. No. 0.0028;
  Matches  144;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              || | ||||  || | || |||||  |  ||    || ||||     |  ||||  |
Db       1547 GCCACGGCGTGCTTCACGCCGAGGGCGGCCAGGTGGTCGAGGAGCAGCCCGGCCTGTTCG 1488

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
              |    | || |  || |  |||   |  ||     ||| |   ||  |          ||
Db       1487 TGGCGCTCGTCCGCGTCCAGGGACAGGTGGGAGCTGAGTACGCCGATGCGGGCGCCGCCG 1428
```

```
Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              ||||  | | |||| ||       |||||||    ||  || ||| | ||| | | |||
Db       1427 ATCCGGACCACGGCGGCGGCGATGCCCCGGCGGTGCCGGCCGGGGGTGAGCGGGAGCAGC 1368

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              | ||          ||| ||| |||  || | |    |       |   |||||| ||  |
Db       1367 ACGTCCTCGGTGCGCTCCACGGTGGCCCGCAGGGAGCAGAGCAGCGCGGGGCCGGCGGCG 1308

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGG 288
              |  ||  |    |    ||||||| || || ||  || | ||||| ||
Db       1307 GTTCCGCCGCCGGAGAGCAGGACCAGATCGCCGGCGGCGGCGAGCCGG 1260


RESULT 8
US-10-297-056-10/c
; Sequence 10, Application US/10297056
; Publication No. US20060030045A1
; GENERAL INFORMATION:
;  APPLICANT: Brett P. Monia
;  APPLICANT:  Madeline M. Butler
;  APPLICANT:  Jacqueline Wyatt
;  APPLICANT:  Isis Pharmaceuticals, Inc.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF C/EBP BETA EXPRESSION
;  FILE REFERENCE: RTSP-0150
;  CURRENT APPLICATION NUMBER: US/10/297,056
;  CURRENT FILING DATE:  2002-12-02
;  PRIOR APPLICATION NUMBER: 09/593,711
;  PRIOR FILING DATE: 2000-06-13
;  NUMBER OF SEQ ID NOS: 244
; SEQ ID NO 10
;   LENGTH: 1500
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (108)...(998)
US-10-297-056-10

  Query Match             19.4%;  Score 57.2;  DB 9;  Length 1500;
  Best Local Similarity   52.4%;  Pred. No. 0.0038;
  Matches  150;  Conservative    0;  Mismatches  133;  Indels    3;  Gaps    1;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              ||| | | | || ||| | |||  || | |    |     ||| |  |  |||| || ||
Db        522 CGAAGCCCGGCTCCGCCTTGAGCGCCGCGGGAGGCGGCGGCGGGAAGCAGGCGGGCGGCG 463

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
                | |  |||| |  ||||| ||   |||| | ||         | |||||  ||  |   |
Db        462 CGGCCTTGGCGCCCGCGCGGCCG---AGGCTCACGTAACCGTAGTCGGCCGGCTTCTTGC 406

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
              || ||| |  | ||| |   |  ||   | |  ||     |  | ||||       | |||
Db        405 TCGGCTTGGCGCCGTAGTCGTCGGCGAAGAGGTCGGAGAGGAAGTCGTGGTGCGCGGGCG 346

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |    |       |||| ||| ||  |   |||||  |   |  ||     |  ||||||||
Db        345 CGGGCGCGGCGAAGTCCGCGGCGGGCGCGAGCGGCTCCAGGTAGGGGCTGAAGTCGATGG 286

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAG 287
              || | | ||| |  | | | ||||  || ||  |||| || || |
Db        285 CGCGCTCGTGCTCGCCAATGGCCGGCTCGGCGGCGGGGGCGCGCGG 240


RESULT 9
US-11-121-086-5
; Sequence 5, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
;   LENGTH: 153376
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-5

  Query Match             19.3%;  Score 56.8;  DB 14;  Length 153376;
  Best Local Similarity   49.7%;  Pred. No. 0.0009;
  Matches  145;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              | ||  | |||| |     |  |||||  || ||  | | ||||||  ||      |  ||
Db     100578 CCAGACCCGTCTCCAGACCTGGGTCCCGCGCCCCCCGGGGCGCGCCCGCCCTTTGTTGCG 100637

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
                ||| ||||      |   |  ||| ||||||  | ||  ||| ||  |  |||||
Db     100638 GGCTCCGGGCGCTTATGAATGGCTGCGGGCGCGGGCGGGGCGGGGGCGGGCGCGGAGAGG 100697

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
              |   |   | |  || |  || |||   |  |||    |||   | |   | |    |
Db     100698 CGCCAGCCCCGCCCCCGCTGCGCTCCCCTTGGTCGCCTGGGCGCTGGGGGCGGGGCGAGG 100757

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |||  |   ||  || || |||  || | ||  ||||| |   |  | |||      |
Db     100758 GGGCGGGCCCGGCTGCCACGTGGGGCGGTGCCGCGGCGGCTCTCATAGGCCCGGGCTCGC 100817

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
              |||| ||  |    ||   |  |||    | |  || |||  || || | ||
Db     100818 GGCCGACTGGGCGCTCGGCGCGCGGCGATTGGCCGGCGCCCGGCCGGCGGAG 100869


RESULT 10
US-11-121-086-6
; Sequence 6, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6
;   LENGTH: 172543
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-6

  Query Match             19.3%;  Score 56.8;  DB 14;  Length 172543;
  Best Local Similarity   49.7%;  Pred. No. 0.00086;
  Matches  145;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              | ||  | |||| |     |  |||||  || ||  | | ||||||  ||      |  ||
Db       4004 CCAGACCCGTCTCCAGACCTGGGTCCCGCGCCCCCCGGGGCGCGCCCGCCCTTTGTTGCG 4063

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
                ||| ||||      |   |  ||| ||||||  | ||  ||| ||  |  |||||
Db       4064 GGCTCCGGGCGCTTATGAATGGCTGCGGGCGCGGGCGGGGCGGGGGCGGGCGCGGAGAGG 4123

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
              |   |   | |  || |  || |||   |  |||    |||   | |   | |    |
Db       4124 CGCCAGCCCCGCCCCCGCTGCGCTCCCCTTGGTCGCCTGGGCGCTGGGGGCGGGGCGAGG 4183
```

```
Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              ||| |   ||  || || |||  || | ||  ||||| |   |  | |||      |
Db       4184 GGGCGGGCCCGGCTGCCACGTGGGGCGGTGCCGCGGCGGCTCTCATAGGCCCGGGCTCGC 4243

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
              |||| ||  |    ||   |  |||    | |  || ||| || || | ||
Db       4244 GGCCGACTGGGCGCTCGGCGCGCGGCGATTGGCCGGCGCCCGGCCGGCGGAG 4295


RESULT 11
US-10-496-351-4
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 9975
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-4

  Query Match             19.1%;  Score 56.4;  DB 7;  Length 9975;
  Best Local Similarity   49.7%;  Pred. No. 0.0028;
  Matches  144;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              | ||||    |||  | ||      ||   ||| ||||  | ||| | ||||| |
Db       6301 GCCGGTTCCGGCGCGGTGGGAGCACGCGGGGCGGAGCCGGTGGGTGTCGCGCGGGCCGCG 6360

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
               ||    ||   |||  | |  ||     |  || | |||||   | | ||  | ||
Db       6361 GGCGCGCGCGAGCCGGAGGCGGGGGGCGCGGTCGATCCCGGCGGGGTGTGTCCGGCGCCT 6420

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              | ||||| ||    |||| |  |  || ||   ||||  | || ||  | ||    |
Db       6421 GCGGTGGGCGCCGCGACGCGCGCACCCGACCCGGAGGCGGCTGCGCCCGGCGTGCCGGGC 6480

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              |       |  | |||||   ||| |  |||  |||  | | ||||||| | ||
Db       6481 CCGGCGGCGCGGCCGGTGGGCGCCGGGCCGGACGGGTCCGCGCGGACCGTCGACGTGGTG 6540

Qy        246 CACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |  |  |   | |     | || | |||| | ||| ||   |   ||||
Db       6541 GGCGCGTCCGGGCCGCTGGTGCCCCTGGTGGTGTCCGCGCGCGACGAGCG 6590


RESULT 12
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
;   LENGTH: 82746
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

  Query Match             19.1%;  Score 56.4;  DB 7;  Length 82746;
  Best Local Similarity   49.7%;  Pred. No. 0.0013;
  Matches  144;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              | ||||    |||  | ||      ||   ||| ||||  | ||| | ||||| |
Db      60900 GCCGGTTCCGGCGCGGTGGGAGCACGCGGGGCGGAGCCGGTGGGTGTCGCGCGGGCCGCG 60841

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
               ||    ||   |||  | |  ||     |  || | |||||   | | ||  | ||
Db      60840 GGCGCGCGCGAGCCGGAGGCGGGGGGCGCGGTCGATCCCGGCGGGGTGTGTCCGGCGCCT 60781

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              | ||||| ||    |||| |  |  || ||   ||||  | || ||  | ||    |
Db      60780 GCGGTGGGCGCCGCGACGCGCGCACCCGACCCGGAGGCGGCTGCGCCCGGCGTGCCGGGC 60721

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              |       |  | |||||   ||| |  |||  |||  | | ||||||| | ||
Db      60720 CCGGCGGCGCGGCCGGTGGGCGCCGGGCCGGACGGGTCCGCGCGGACCGTCGACGTGGTG 60661

Qy        246 CACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |  |  |   | |     | || | |||| | ||| ||   |   ||||
Db      60660 GGCGCGTCCGGGCCGCTGGTGCCCCTGGTGGTGTCCGCGCGCGACGAGCG 60611


RESULT 13
US-10-301-480-566618
; Sequence 566618, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 566618
;   LENGTH: 663
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-566618

  Query Match             18.8%;  Score 55.4;  DB 10;  Length 663;
  Best Local Similarity   53.5%;  Pred. No. 0.011;
  Matches  116;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         26 CCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGT 85
              || ||  || |||  |||  || || || |||        ||  || ||| |  | ||
Db         15 CCGCCAGCCCGCCCGGCGGCCGCGCAGCCCGCACTCGTCGCCGCGCCTCGGAGTTTGAGC 74

Qy         86 GCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCG 145
              | ||||    |||||||  |   | | |||  | |   |   ||| |||  |  | |||
Db         75 CCCAGCGCTGCCGGAGCGCAGGCACGCCCGGGGCGCGGGGTCGGAGCGCGCAGCGCGGCG 134

Qy        146 CCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGG 205
              ||   | |  || ||||     |      |  || || ||  |  | |||||      || ||
Db        135 CCGCCCCCCGGCCCTCGGCCCCCCAACCCCGGCGCCCCCGGAGCGGCGCGCGGAGGGAGG 194

Qy        206 CGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGG 242
              ||||||  | |||  || || | ||||| ||   ||
Db        195 AGGGCGGGCGGGCGCGGCGGGCCGGGCCGGCGGGCGG 231
```

```
RESULT 14
US-10-301-480-1180027
; Sequence 1180027, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1180027
;   LENGTH: 663
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1180027

  Query Match             18.8%;  Score 55.4;  DB 10;  Length 663;
  Best Local Similarity   53.5%;  Pred. No. 0.011;
  Matches  116;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         26 CCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGT 85
              || ||  || |||  |||  || || || |||        ||  || |||  |  | |
Db         15 CCGCCAGCCCGCCCGGCGGCCGCGCAGCCCGCACTCGTCGCCGCGCCTCGGAGTTTGAGC 74

Qy         86 GCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCG 145
               |  ||||    |||||||  |   | | |||  | |   |   ||| |||  |  | |||
Db         75 CCCAGCGCTGCCGGAGCGCAGGCACGCCCGGGGCGCGGGGTCGGAGCGCGCAGCGCGGCG 134

Qy        146 CCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGG 205
              ||   | |  || ||||    |      |  || || ||  |  | |||||     || ||
Db        135 CCGCCCCCCGGCCCTCGGCCCCCCAACCCCGGCGCCCCCGGAGCGGCGCGCGGAGGGAGG 194

Qy        206 CGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGG 242
               ||||||  | |||  || || | ||||| ||   ||
Db        195 AGGGCGGGCGGGCGCGGCGGGCCGGGCCGGCGGGCGG 231


RESULT 15
US-10-496-351-1/c
; Sequence 1, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 14055
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-1

  Query Match             18.8%;  Score 55.4;  DB 7;  Length 14055;
  Best Local Similarity   50.6%;  Pred. No. 0.0039;
  Matches  134;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              || || ||   | ||||| ||||  |  ||| |||| ||   |||    || ||   |||
Db       4544 GCCAGCGCCTCCAGCGCGGCGAGCGCGCCCGCCCGGTCGGCGGCGACCACGACGGCGCGC 4485

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
              |       | || | | || |||   |||      |    ||| ||||||||       |
Db       4484 TCGCCGAGGTCGGCCCGGCCGGCTCTCAGCGCGCCCGCCAGCTCGGCGGCCGTGACGTCG 4425

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
                | ||||| |||| | ||  |||| |  |       | ||   ||||| ||
Db       4424 GAGGGGGCGAGGAAGTCGGCCAGCCGCCGGGCCTGGGCGCGCAGCGCCGCCGCGTCGACG 4365

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||      | ||     | ||||  || ||   |||||||||   |    |   |||  |
Db       4364 GCGGACACCGGCAGCGGCACGGGCCCGCGCTCGGGCGCGGCGGGCGCGGCGGTCTCGGCG 4305

Qy        241 GGGCCCACCTGCGCTTCCAGGACCG 265
              ||  ||  || |    ||| |  ||
Db       4304 GGCGCCTGCTCCAGGACCACGTGCG 4280


Search completed: April 27, 2006, 15:53:03
Job time : 186.244 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 12:17:22 ; Search time 295.759 Seconds
(without alignments)
8248.165 Million cell updates/sec

Title: US-10-729-802-1_COPY_77000_77294
Perfect score: 295
Sequence: 1 gcgagggcggtctgcgcgtc.....ggggccgagcaggtgcagcg 295

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	295	100.0	77294	8	US-10-729-802-1	Sequence 1, Appli
	2	66	22.4	138203	9	US-10-819-386A-1	Sequence 1, Appli
c	3	65.2	22.1	616	7	US-10-767-701-3163	Sequence 3163, Ap
c	4	64.2	21.8	653	7	US-10-767-701-4983	Sequence 4983, Ap
	5	63.6	21.6	763	7	US-10-437-963-49370	Sequence 49370, A
c	6	63.4	21.5	30000	3	US-09-980-217-2	Sequence 2, Appli
c	7	63.2	21.4	1180	7	US-10-437-963-24598	Sequence 24598, A
c	8	63.2	21.4	1614	7	US-10-437-963-81245	Sequence 81245, A
c	9	63	21.4	1235	7	US-10-437-963-14082	Sequence 14082, A
	10	62.4	21.2	71989	3	US-09-727-889-2	Sequence 2, Appli
	11	62.4	21.2	71989	8	US-10-849-462-2	Sequence 2, Appli
	12	62	21.0	972	8	US-10-425-115-17653	Sequence 17653, A
	13	62	21.0	1436	8	US-10-425-115-104721	Sequence 104721,
c	14	61.8	20.9	2184	7	US-10-282-122A-14290	Sequence 14290, A
	15	61.6	20.9	1488	7	US-10-437-963-42386	Sequence 42386, A
c	16	61.4	20.8	446	7	US-10-767-701-1943	Sequence 1943, Ap
c	17	61.2	20.7	1683	3	US-09-758-759-148	Sequence 148, App
c	18	61.2	20.7	1683	10	US-11-021-825-148	Sequence 148, App
c	19	61.2	20.7	109519	3	US-09-758-759-1	Sequence 1, Appli
c	20	61.2	20.7	109519	10	US-11-021-825-1	Sequence 1, Appli
c	21	60.8	20.6	4257	3	US-09-825-288A-1	Sequence 1, Appli
c	22	60.6	20.5	661	7	US-10-767-701-9110	Sequence 9110, Ap
c	23	60.4	20.5	763	7	US-10-437-963-49370	Sequence 49370, A
c	24	60.2	20.4	753	6	US-10-156-761-6575	Sequence 6575, Ap
c	25	60.2	20.4	804	7	US-10-437-963-99082	Sequence 99082, A
	26	60.2	20.4	9025608	6	US-10-156-761-1	Sequence 1, Appli
c	27	60	20.3	4667	8	US-10-723-860-5759	Sequence 5759, Ap
	28	59.8	20.3	1162	3	US-09-969-708-571	Sequence 571, App
	29	59.8	20.3	1162	9	US-10-843-641A-8042	Sequence 8042, Ap
	30	59.8	20.3	2325	8	US-10-723-860-6692	Sequence 6692, Ap
	31	59.8	20.3	3774	8	US-10-723-860-2538	Sequence 2538, Ap
	32	59.8	20.3	3774	9	US-10-756-149-2350	Sequence 2350, Ap
	33	59.8	20.3	3774	9	US-10-840-060-226	Sequence 226, App
c	34	59.6	20.2	1212	6	US-10-156-761-4023	Sequence 4023, Ap
c	35	59.6	20.2	1239	7	US-10-282-122A-13867	Sequence 13867, A
c	36	59.6	20.2	1488	7	US-10-437-963-97166	Sequence 97166, A
c	37	59.6	20.2	3957	6	US-10-200-562-193	Sequence 193, App
c	38	59.6	20.2	3957	6	US-10-237-551-193	Sequence 193, App
c	39	59.6	20.2	3957	9	US-10-945-050-193	Sequence 193, App
	40	59.6	20.2	154746	3	US-09-827-688-8	Sequence 8, Appli
c	41	59.6	20.2	154746	3	US-09-827-688-8	Sequence 8, Appli
c	42	59.4	20.1	1914	6	US-10-156-761-5963	Sequence 5963, Ap
c	43	59.2	20.1	1232	7	US-10-437-963-17794	Sequence 17794, A
c	44	59.2	20.1	1677	7	US-10-437-963-43981	Sequence 43981, A
c	45	59	20.0	1252	7	US-10-437-963-71363	Sequence 71363, A

ALIGNMENTS

RESULT 1
US-10-729-802-1
; Sequence 1, Application US/10729802
; Publication No. US20050032184A1
; GENERAL INFORMATION:
; APPLICANT: Julien, Bryan
; APPLICANT: Reid, Ralph C.
; TITLE OF INVENTION: Disorazole Polyketide Synthase Encoding Polynucleotides
; FILE REFERENCE: 020547-003560US
; CURRENT APPLICATION NUMBER: US/10/729,802
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 60/512,892
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: US 60/484,934
; PRIOR FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: US 60/473,311
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/465,038
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US 60/455,521
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/431,272
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 77294
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-729-802-1

Query Match 100.0%; Score 295; DB 8; Length 77294;
Best Local Similarity 100.0%; Pred. No. 9.1e-61;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77000 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 77059

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77060 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 77119

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77120 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 77179
```

```
Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77180 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 77239

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77240 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 77294


RESULT 2
US-10-819-386A-1
; Sequence 1, Application US/10819386A
; Publication No. US20050089884A1
; GENERAL INFORMATION:
;  APPLICANT: Korea advanced Institute of Science and Technology
;  APPLICANT:  Shanghai Jiaotong University
;  APPLICANT:  LEE, Sang Yup
;  APPLICANT:  DENG, ZIXIN
;  APPLICANT:  CHEN, SHI
;  APPLICANT:  JEONG, Ki Jun
;  APPLICANT:  ZHOU, XIUFEN
;  TITLE OF INVENTION: Genes for the Synthesis of FR-008 Polyketides
;  FILE REFERENCE: PO03-B015
;  CURRENT APPLICATION NUMBER: US/10/819,386A
;  CURRENT FILING DATE:  2004-04-06
;  PRIOR APPLICATION NUMBER: KR10-2003-0074035
;  PRIOR FILING DATE: 2003-10-23
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 138203
;   TYPE: DNA
;   ORGANISM: Streptomyces sp. FR-008
US-10-819-386A-1

  Query Match             22.4%;  Score 66;  DB 9;  Length 138203;
  Best Local Similarity   51.7%;  Pred. No. 1.6e-07;
  Matches  150;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              ||||||   |    ||   | | |   | || ||||||  |  | ||||| || ||
Db     126896 CGAGGGTCTTGGCGGCCATGGCGGCGGCGACCTGGTCGAGCGGGGTGTCGGCGAGCGGCG 126955

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
                | |  | || |||||  |||||||| |     ||  ||| | |||| ||   ||| ||
Db     126956 CGGCCTGGCCGACGCCGGCGGCGTGCACGACGCCGGTGAGGGGGGCGTCCTCGGGCACCG 127015

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
                  || |||| |||      | |||    |||  |||| |||      |  ||
Db     127016 CGGCGAGGACGGCGGCGAGGGCGTCGCGGTCGGCGCTGTCGCAGGCGGTGAGGGTGACGC 127075

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
                ||  |||  | |||     |||| |  |    ||||||||||    | ||  |  |
Db     127076 GGGCGCCCAGTTCCTCCAGTTCGGCGCGGAGTTCGGCGGCGCCGGGGGCGTCGGGGCCGC 127135

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGC 291
              |||   |        | |||   |   || ||  | || |||||  | ||
Db     127136 GGCGGCCGGTGAGGACGAGGTGGGCGGCGCCGGAGCGGGCGAGCCAGCGC 127185


RESULT 3
US-10-767-701-3163/c
; Sequence 3163, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 3163
;   LENGTH: 616
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS53158_1
US-10-767-701-3163

  Query Match             22.1%;  Score 65.2;  DB 7;  Length 616;
  Best Local Similarity   52.6%;  Pred. No. 1.6e-06;
  Matches  142;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              || ||| | |  |  | | | |  |  | |      |   | |  || |||  || |  |
Db        342 GGTGGTGTTCATGGTGCGCTTCCGCAGCAGCGGCTGGCGGGGCCCGCCGCCGCGGTCGGG 283

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
               ||| |||||| | ||  | |||||  |||| ||      | ||||  |||||  |  | | |
Db        282 GCCGCCGTCGCGGAGGCGGGGCAGGGCCGCGGGGGCGCGGCCGGCGCGCGGGCGGCGCGG 223

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              |  | ||||| ||         | | |   | ||   ||| | |    || || | ||  |
Db        222 CGGGGGGGCGGCGCCGGCGCCGCCGTCGATGCAGCCGCCGTGGCAGTCGGGGTGGGCGCC 163

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
                | | |||||     ||||| |||||||  ||| || ||||     ||| || |  | |  |
Db        162 CGGCGGCGCGGGGCGGGGGGGGGGCGGCCGCGGGGGGCCGCCGGGGCGGTGGCGGCGCAC 103

Qy        246 CACCTGCGCTTCCAGGACCGGACCGTCGTG 275
              | ||     |  | |||  | |  | |
Db        102 AGGCCGCCGCGGCGCGCCCGCCCTGCTGCG 73


RESULT 4
US-10-767-701-4983/c
; Sequence 4983, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 4983
;   LENGTH: 653
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS95688_1
US-10-767-701-4983

  Query Match             21.8%;  Score 64.2;  DB 7;  Length 653;
  Best Local Similarity   51.2%;  Pred. No. 2.7e-06;
  Matches  150;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              || ||||||   ||||| ||     |    | ||  ||| ||||  ||||  | ||| ||
Db        318 CGTGGGCGGGGGGCGCGACGGCCCCGCGGGCCCCCCCGTGCGCCGGGGCCCCCCGCGGCC 259

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
                |   |  || | |||||     ||       ||  |    |  || | | |||  |   ||
Db        258 GCCGGCGCGCCTGGCCGCCCCGGTCGGCCGGCCGGCACGGGCCGACCGGCCGGCGGGGCC 199

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
                | |   |||||||  |  |   | | | ||   | ||||   |  | | |||      |
Db        198 CCAGTGGGAGGGCGGGGCGCGTGGGCGGCCGGGGGTGCTCCAAGCGGGGGGCGCCCCGAA 139
```

```
Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |||  | |||    |  ||||||||||| ||||||  |  |      | | ||  ||
Db        138 GGAGGGAGGGGGCGCAGACCTGGGCGGGCGGCCGCGCGGGACGCGCGGCCCGGCCGGGGG 79

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGC 294
              ||| |  | | |   || |  | ||       |  ||||  | ||||| |  ||
Db         78 GGCGCCACCGGGAAACCCGCCCAGGTGGCGGGACGGGGAAGGGCAGGGGTGGC 26


RESULT 5
US-10-437-963-49370
; Sequence 49370, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 49370
;   LENGTH: 763
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(763)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_51959C.1
US-10-437-963-49370

  Query Match             21.6%;  Score 63.6;  DB 7;  Length 763;
  Best Local Similarity   52.1%;  Pred. No. 3.6e-06;
  Matches  135;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              ||| | | || |  || | |  |  ||  ||    | ||| |  ||||| ||||   |
Db        306 GGCCGCCCGCCCCCCGCGCCCNCCGCCCCCGCGCCCCCCGCCGCGGGCGCCCCGCGCGGC 365

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
                ||  |  ||||  |  || |  ||||      ||| |  |||| |||| ||| |||
Db        366 GGGCGCCGCCGCGCGNCCCCGCGGCGGCGCCGCCCCCGCCGGCGCCCCCGGCGCCCGCCG 425

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              |     ||        |  ||||| || | ||||  | |  |||   | | || | ||| |
Db        426 CGGCCCGGGCCGCGCCGCCGCCCCGCCCCCGCGGGCCCCCGCGGNGGGGGGGCGGGCCGC 485

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              || |    |  || | |  || ||||||  |||||||| | | |   |   |   |    |
Db        486 GGCCCCGCCCCCCGCCGCCGCCGGCGGCCCCGCGGCGNCNGCCGCCCGCCCCCCGCCCGC 545

Qy        246 CACCTGCGCTTCCAGGACC 264
              | ||  | |   | || ||
Db        546 CGCCCCCCCCCGCCGGCCC 564


RESULT 6
US-09-980-217-2/c
; Sequence 2, Application US/09980217
; Publication No. US20040219645A1
; GENERAL INFORMATION:
;  APPLICANT: Biotica Technology Limited
;  APPLICANT:  Leadlay, Peter F
;  APPLICANT:  Staunton, James
;  APPLICANT:  Oliynyk, Marko
;  TITLE OF INVENTION: Polyketides and their synthesis
;  FILE REFERENCE: IS/BP5858469
;  CURRENT APPLICATION NUMBER: US/09/980,217
;  CURRENT FILING DATE:  2002-05-06
;  PRIOR APPLICATION NUMBER: PCT/GB00/02072
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: GB 9912563.5
;  PRIOR FILING DATE: 1999-05-28
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 30000
;   TYPE: DNA
;   ORGANISM: Streptomyces cinnamonensis
;   FEATURE:
;   OTHER INFORMATION: Nucleotides 30001 - 60000 of the monensin
;   OTHER INFORMATION: biosynthetic gene cluster
US-09-980-217-2

  Query Match             21.5%;  Score 63.4;  DB 3;  Length 30000;
  Best Local Similarity   52.5%;  Pred. No. 1.1e-06;
  Matches  139;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          6 GGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGT 65
              | | | | | ||| ||  | || | | |  | || |  || |  | |  || |   ||
Db      11832 GCCCGCCGGGGCGACGCTGCCCGCGGCCGCGGTGCGCCGGGCGGTGCGGGCCCCCGGGGT 11773

Qy         66 CCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCG 125
              | |  |  || |  ||||| ||||| ||||| |  | |  | |||  || || |
Db      11772 CGCACCCGCGGCCAGGGCGCGCAGGAGCGCGGGCCGGGACGGGGCGGGCTGCGCGGCGAC 11713

Qy        126 CTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTC 185
              || | |||||||  | | ||  |  |||| ||  |||| |||| ||| ||   || ||
Db      11712 GGCGCGCGCGTCGAGGTCCACCGCGACCACGTGCGGGCGGCCGTGGGCGCGGGCGGCGTC 11653

Qy        186 GGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCC 245
              |   || ||      |      |||  | | ||| || |         ||  |  ||| |  |
Db      11652 GAGGAGGGCCAAGCCACCTTCGGTGGTCAGCGGGGTGAACCCGGTGCGCTTCATGCGTGC 11593

Qy        246 CACCTGCGCTTCCAGGACCGGACCG 270
              ||  | ||  ||   ||    | |||
Db      11592 CAGGTCCGTCTCGCCGAGGTGTCCG 11568


RESULT 7
US-10-437-963-24598/c
; Sequence 24598, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 24598
;   LENGTH: 1180
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1180)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
```

```
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_29567C.1
US-10-437-963-24598

  Query Match            21.4%;  Score 63.2;  DB 7;  Length 1180;
  Best Local Similarity   52.6%;  Pred. No. 3.8e-06;
  Matches  131;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy         37 CCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCG 96
              | | ||| |||||||    | |||  || | |||||  | || | | |    | ||| |
Db        565 CGGGGCGGGCGGGCGCGCGGGCGCCGCGGCGCGGCGGNGGCGGGCGGGGGGGCGGGCGGG 506

Qy         97 AGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACG 156
               ||  || ||| || |  | | ||| |     ||  |  | ||  | |  |  ||||   |
Db        505 CGGGCCGGGGCCGCGCCGGCCGCCCGCGCGCCGGCGCGGCCGGCCCGGCCGGGGCGGCGG 446

Qy        157 CGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGC 216
              ||  | |  | | ||| | |  ||||  ||  | ||  |  | | | |||| ||||||
Db        445 CGGGCCCCCGGGCGCGGGCGCGGCGCGGCGCCCGGCCGGGGCGCCGCGGCGCGCGGCGCG 386

Qy        217 GCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGG 276
              |  | ||       |||| |   |||  |    |||| |     | |  | | ||  |
Db        385 GGNGGGCGCCNGGCGCCGCGGGCCGGGGGCGGGGGCGCCCCGCNCCCGGCGCGGGCGCCG 326

Qy        277 GGGCCGAGC 285
              |  ||| ||
Db        325 GNCCCGCGC 317


RESULT 8
US-10-437-963-81245/c
; Sequence 81245, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 81245
;   LENGTH: 1614
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_80791C.1
US-10-437-963-81245

  Query Match            21.4%;  Score 63.2;  DB 7;  Length 1614;
  Best Local Similarity   51.8%;  Pred. No. 3.4e-06;
  Matches  143;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy         20 CGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGC 79
              ||  | | || |    ||  |||||||    || |  |||||  | |         |||
Db       1012 CGTCGGCGTCGGCGAGGGTGAGCGCGCTCACGCCCTGCCGCGCCTTGAGCCGCGCGCGCT 953

Qy         80 GGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGA 139
                |  | | | |||     |   |  || |  | |  | | | | | ||  |
Db        952 CCGACGGTGGCAGGGCGTCCCACTGCTCCCGCCACTCGCACACCATCACCACGCCGGTGG 893

Qy        140 TCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCT 199
              || || |  ||    |   |   | ||||  |  ||||  ||||    || ||  |
Db        892 CCTCCGTCATGCCGTAGGCGTGCGTGACGTGGAAGCCGAGCCGCTCGACGCGCTCGAGCA 833

Qy        200 CGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCA 259
                |  |||||||||||||||| |||||||  ||  | || | |||||    | |||  |||
Db        832 GCGCCGCGGGCGGCGGCGCGCCGCCGGTGAGCACCTCCACGGGGCGGCGCAGCGGCTCCC 773

Qy        260 GGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              ||  ||  |||||| || ||  |   | | |||  ||
Db        772 GGCACGCGCCGTCGAGGAGGATGTTGAAGAGCACCG 737


RESULT 9
US-10-437-963-14082/c
; Sequence 14082, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 14082
;   LENGTH: 1235
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_20052C.1
US-10-437-963-14082

  Query Match            21.4%;  Score 63;  DB 7;  Length 1235;
  Best Local Similarity   51.6%;  Pred. No. 4.2e-06;
  Matches  144;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy          8 CGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCC 67
              |||    |||  || |  ||  || |||||| | |||  ||        | ||||   | ||
Db       1167 CGGCGCCCGCCGCGCGCCCCGGCGGCCGGCGGCGCGGCCGCCGCCGCCCCCGCGCGGCCC 1108

Qy         68 CGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCT 127
              | ||  |  ||||      | |||||| |  |  | ||||||||   |  ||||  |||
Db       1107 CCGCCCCCCGCGCGCCGCGCCCGGCGCGGGCCGGCCCCGGCGGCCGCGGCGCCCCCGCGCG 1048

Qy        128 CGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGG 187
              |   | || |   | || |||  ||   | |   | || || | |  ||| |||    ||
Db       1047 CCCCGCCGCCCCCCCGCCCCCCCCGGGGGGGTGGGGGGTCGGGGGGCTGGGGCGGGGGGG 988

Qy        188 ACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCA 247
                | | |      ||||| |||||| || | || |  ||   ||  |  |  ||||  |
Db        987 GGGGGGGGGTGCGGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCC 928

Qy        248 CCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCA 286
              |   | |  | |  |   |   | ||  | |   |||||
Db        927 CGCCCCCCCCAAAAAAAAGGGTGACGGAGTGCGGGAGCA 889


RESULT 10
US-09-727-889-2
; Sequence 2, Application US/09727889
; Publication No. US20010051506A1
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
```

```
;  CURRENT APPLICATION NUMBER: US/09/727,889
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: 09/443,501
;  PRIOR FILING DATE: 1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-727-889-2

  Query Match             21.2%;  Score 62.4;  DB 3;  Length 71989;
  Best Local Similarity   51.4%;  Pred. No. 1.4e-06;
  Matches  144;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         14 GCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGT 73
              ||||||| || |  |    |||| |||| || |  ||||      || || ||  |||
Db      65289 GCGCGTCGCGGCCGGCGTCGCGGCGGAGCTCGTGCACGGCCTCGGCCTGCGCCCGCGCGG 65348

Qy         74 CGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGG 133
                | ||| ||   | ||   || | |||||    ||  |  ||           | |
Db      65349 GATCACGGACGCCGACGCCGCCGCGCTGCGCGCGCTCCGCCCCCCCGGACGGCGAGGGGCA 65408

Qy        134 CGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCG 193
              || ||    |   || ||||| ||||| ||    || |||||  || | ||  |||
Db      65409 CGGCGCCTTCGTCCGGGCGCTCGCGCGCGATCTCGCGCGCGCGGGGGGGCGCCGGCGTCGC 65468

Qy        194 CGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCG 253
              |||| |||| | ||| | || || |  || |        ||   | |    | ||
Db      65469 CGTCGTCGGCGACGGCCAGCCGCCCATCGTCCACGCCCTCGGGCACGTCATCAACGCCGC 65528

Qy        254 CTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
                |||    |||| | | |  | || ||| |  |||| |
Db      65529 GCTCCGCAGCCGGGCGGCCTGGATGGTCGATCCTGTGCTG 65568


RESULT 11
US-10-849-462-2
; Sequence 2, Application US/10849462
; Publication No. US20040253697A1
; GENERAL INFORMATION:
;  APPLICANT: Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: RECOMBINANT METHODS AND MATERIALS FOR
;  TITLE OF INVENTION:  PRODUCING EPOTHILONE AND EPOTHILONE DERIVATIVES
;  FILE REFERENCE: 300622003101
;  CURRENT APPLICATION NUMBER: US/10/849,462
;  CURRENT FILING DATE:  2004-05-18
;  PRIOR APPLICATION NUMBER: US/09/724,878
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: US 09/443,501
;  PRIOR FILING DATE: 1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Construct
US-10-849-462-2

  Query Match             21.2%;  Score 62.4;  DB 8;  Length 71989;
  Best Local Similarity   51.4%;  Pred. No. 1.4e-06;
  Matches  144;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         14 GCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGT 73
              ||||||| || |  |    |||| |||| || |  ||||      || || ||  |||
Db      65289 GCGCGTCGCGGCCGGCGTCGCGGCGGAGCTCGTGCACGGCCTCGGCCTGCGCCCGCGCGG 65348

Qy         74 CGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGG 133
                | ||| ||   | ||   || | |||||    ||  |  ||           | |
Db      65349 GATCACGGACGCCGACGCCGCCGCGCTGCGCGCGCTCCGCCCCCCCGGACGGCGAGGGGCA 65408

Qy        134 CGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCG 193
              || ||    |   || ||||| ||||| ||    || |||||  || | ||  |||
Db      65409 CGGCGCCTTCGTCCGGGCGCTCGCGCGCGATCTCGCGCGCGCGGGGGGGCGCCGGCGTCGC 65468

Qy        194 CGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTGCG 253
              |||| |||| | ||| | || || |  || |        ||   | |    | ||
Db      65469 CGTCGTCGGCGACGGCCAGCCGCCCATCGTCCACGCCCTCGGGCACGTCATCAACGCCGC 65528

Qy        254 CTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAG 293
                |||    |||| | | |  | || ||| |  |||| |
Db      65529 GCTCCGCAGCCGGGCGGCCTGGATGGTCGATCCTGTGCTG 65568


RESULT 12
US-10-425-115-17653
; Sequence 17653, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 17653
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(972)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_116102C.1
US-10-425-115-17653

  Query Match             21.0%;  Score 62;  DB 8;  Length 972;
  Best Local Similarity   50.5%;  Pred. No. 7.8e-06;
  Matches  149;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              | |  || ||   | |||  | ||          ||| || |||||||  |  | |  |||
Db        424 GGGGNGGGGGGGGGGGGCGGGGGGGGGGGGGCGGCGGGCGCGCGCGCGCCCTCCCCCGCGC 483

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                || | | | | ||||||||||| |   |     |     | |||  ||| || |||| |
```

```
Db        484 CCCGCCGCCGGGGCGCCGCGGGCGCGGCCGGCGCCCTCCGGGGCGCGGGCGCGGGGCCGC 543

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180

Db        544 GGCGGCGCGGCGGCGCCGGGCCGCGGGCGGCGGCGCCGGCCTCGCCCGCGCGCTCGCGCG 603

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240

Db        604 CGGGCGCCGGGTCGGGGCGGGGCGGGGCGCGGGGGGGGGGGGGGGCGGGGGGGCGCGG 663

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295

Db        664 TGGGCGGCGGGGGGGGGCGGGGGGGGCGGGGCGGCGGGGGGGGGGCGGGGGCG 718


RESULT 13
US-10-425-115-104721
; Sequence 104721, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 104721
;   LENGTH: 1436
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_27008C.1
US-10-425-115-104721

  Query Match             21.0%;  Score 62;  DB 8;  Length 1436;
  Best Local Similarity   52.8%;  Pred. No. 6.8e-06;
  Matches  134;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy         41 GCGCGCGGGCGGCGCGCCGCTTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGA 100

Db        545 GCGTGCGGGGCTGCAGGAGGCTCGGCGGAGGCACGCGCGCGCGGTGGCGGCGGCCGTGCA 604

Qy        101 GCGCGGCGGCCCGCGGCCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCT 160

Db        605 GCGCGCCGCGGCGGGGCGGCTGCGGCTGGCGGAGGCCGAGCTGGGGCGCGCCCGCCGCCG 664

Qy        161 CGCGGGCGTGCGCGTGGAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGG 220

Db        665 CAACGCCGAGCTGGAGGAGAGGCTCCGGCAGCTGGCCGCCGAGGGGCAGGCGTGGCTGGG 724

Qy        221 CGCCGGTCAGGCCGTCGATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGC 280

Db        725 CGTCGCCCGGAGCCACGAGGCCGTCGCCGCGGGGCTCCGCGCCACGCTGGACCAGCTGCT 784

Qy        281 CGAGCAGGTGCAGC 294

Db        785 GCAGCAGCAGCAGC 798


RESULT 14
US-10-282-122A-14290/c
; Sequence 14290, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14290
;   LENGTH: 2184
;   TYPE: DNA
;   ORGANISM: Burkholderia mallei
US-10-282-122A-14290

  Query Match             20.9%;  Score 61.8;  DB 7;  Length 2184;
  Best Local Similarity   52.9%;  Pred. No. 6.5e-06;
  Matches  157;  Conservative    0;  Mismatches  137;  Indels    3;  Gaps    1;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60

Db        683 GCGAGCGCGAGCTGCGCGTCGGCGTCGTCGGGCGCGCGCGCGAGCAGCGGCGCGAGCGCC 624

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120

Db        623 GCGCGCGCGGTGTCGTAGCGCCCCTGCGCGCGCGCGAGCTCGGCCGCCTCGAGCGCATAG 564

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180

Db        563 CGGCGATCGGCCCTGCCTGCCGGGCTCGCGCGATCGAGCGCGCGCGCGGCCGCCTCGTAA 504

Qy        181 GCCTCG---GACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCG 237

Db        503 TCGTCGCGATCGAGCGCGTCATCGGTTTCGCGCAGCACGAGCCGCAGCGCCTGATCCTCG 444

Qy        238 ATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGC 294

Db        443 AGGCGCGCGCTTTGCGCGGCGTCGAGCGGCGCGCCGCCGCGGCGCAGCGTGTCGAGC 387


RESULT 15
US-10-437-963-42386
; Sequence 42386, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
```

```
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 42386
;   LENGTH: 1488
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1488)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_45643C.1
US-10-437-963-42386

  Query Match            20.9%;  Score 61.6;  DB 7;  Length 1488;
  Best Local Similarity  52.3%;  Pred. No. 8.3e-06;
  Matches  136;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              ||  |||| |  ||||| |||||||  ||     ||  |||| ||  |  || |  |||
Db       1047 CGTCGGCGATGAGCGCGACGAGGTCGCCCACGAGGTGGAGCACGGTGTAGGAGTCCCGGC 1106

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
               || |  |||||||  |  | |||    |||   |  | ||||  |||| ||   |||||
Db       1107 CCGGCGGGGCGTCGGAGTCGCCGTAGCCGCGGAGGTCGGGCGCCACGGCGCGGAACCCCC 1166

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
              |    ||||||||   |||||    || | |  |  |  | |  |  |  |  | | |
Db       1167 GCGCGGCGAGGGCGCGCATCTGGTGGCGCCACGAGTACCAGAGCTCCGGGAAGCCGTGGA 1226

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |       ||  |||  | ||    | ||   || | | ||  ||  |||||   |   |
Db       1227 CGAGCAGCACGGCGGGCCCGTCCTCCGGGCCCGCCTCCGCCACGTGCAGGCGCACCCCGC 1286

Qy        242 GGCCCACCTGCGCTTCCAGG 261
                |||||| | |   | ||
Db       1287 TTGCCACCTCCACCGTCCGG 1306


Search completed: April 27, 2006, 14:40:06
Job time : 298.759 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 15:53:17 ; Search time 1462.06 Seconds
(without alignments)
9440.233 Million cell updates/sec

Title: US-10-729-802-1_COPY_77000_77294
Perfect score: 295
Sequence: 1 gcgagggcggtctgcgcgtc.....ggggccgagcaggtgcagcg 295

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	80.6	27.3	932	10	CNS0072Q	AL066742 Drosophil
	2	78.6	26.6	925	10	CNS0091P	AL053013 Drosophil
	3	78.2	26.5	935	10	CNS006XK	AL066051 Drosophil
c	4	77	26.1	925	10	CNS0091P	AL053013 Drosophil
c	5	76.8	26.0	844	10	CNS0052P	AL056652 Drosophil
c	6	75.8	25.7	1755	10	AG057427	AG057427 Pan trogl
	7	75.4	25.6	1134	10	AG043044	AG043044 Pan trogl
c	8	75	25.4	1306	10	AG063529	AG063529 Pan trogl
	9	75	25.4	1444	10	CG769390	CG769390 TcB52.3_G
	10	75	25.4	1552	10	AG430101	AG430101 Mus muscu
	11	74.8	25.4	1009	10	CNS010EW	AL098882 Drosophil
	12	74.6	25.3	931	10	AG076820	AG076820 Pan trogl
c	13	74.4	25.2	932	10	CNS0072Q	AL066742 Drosophil
c	14	74.4	25.2	935	10	CNS006XK	AL066051 Drosophil
	15	74	25.1	1185	5	BQ939301	BQ939301 AGENCOURT
c	16	73.6	24.9	1161	10	CL477996	CL477996 SAIL_283_
	17	73.2	24.8	801	6	CB898588	CB898588 tric015xe
	18	73.2	24.8	801	7	CF868503	CF868503 tric015xe
	19	73	24.7	1821	10	CL090560	CL090560 ISB1-17N1
c	20	72.8	24.7	1299	10	AG039481	AG039481 Pan trogl
	21	72	24.4	1713	10	AG435689	AG435689 Mus muscu
	22	71.4	24.2	1153	11	CNS07BVP	AL438395 T3 end of
c	23	71.2	24.1	1153	11	CNS07BVP	AL438395 T3 end of
c	24	71	24.1	1057	11	CNS046BO	AL276477 Tetraodon
	25	71	24.1	1075	10	AG073983	AG073983 Pan trogl
c	26	70.4	23.9	842	10	AG058791	AG058791 Pan trogl
c	27	70.4	23.9	904	10	AG857798	AG857798 Oryza sat
c	28	70.2	23.8	1124	10	CL491616	CL491616 SAIL_559_
c	29	70	23.7	871	10	CW951452	CW951452 TcB36.4_G
c	30	69.8	23.7	1046	5	BQ643604	BQ643604 AGENCOURT
c	31	69.8	23.7	1242	3	BM911414	BM911414 AGENCOURT
c	32	69.8	23.7	1798	10	AG171124	AG171124 Pan trogl
	33	69.6	23.6	950	5	BQ715722	BQ715722 AGENCOURT
c	34	69.6	23.6	951	10	CL476324	CL476324 SAIL_252b
c	35	69.6	23.6	1009	10	CNS010EW	AL098882 Drosophil
c	36	69.6	23.6	1038	10	AG060254	AG060254 Pan trogl
c	37	69.6	23.6	1878	10	AG393083	AG393083 Mus muscu
c	38	69.4	23.5	776	10	CNS010RY	AL099352 Drosophil
c	39	69.4	23.5	1070	9	BZ557810	BZ557810 pacs1-60_
c	40	69.2	23.5	796	9	AZ192927	AZ192927 SP_1022_B
	41	69.2	23.5	1057	5	BU526464	BU526464 AGENCOURT
c	42	69	23.4	855	5	BU963646	BU963646 AGENCOURT
c	43	69	23.4	1017	10	CL463304	CL463304 SAIL_1182
	44	68.8	23.3	715	10	CW945121	CW945121 TcB27.2_G
	45	68.8	23.3	781	10	CW938741	CW938741 TcB18.4_F

ALIGNMENTS

RESULT 1
CNS0072Q
LOCUS CNS0072Q 932 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066742
VERSION AL066742.1 GI:4945205
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 932)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .932
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14B09"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

```
  Query Match             27.3%;  Score 80.6;  DB 10;  Length 932;
  Best Local Similarity   35.3%;  Pred. No. 4.7e-05;
  Matches  104;  Conservative   77;  Mismatches  114;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              | | |:::: ::|: | | || :: :| : | | ||  | |:  |:||   | |:::||| 
Db        580 GGGGGSBSBTKYTBKGKGGCGGSSGSCCSGGSCRGGGGGCGSCGGSGGCGCGGGSSSCGC 639

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
               :  | :|||   |  ::|: | :: |: |:|:|  |  ::|| |:|||| ::||:|| 
Db        640 SCSCGCGSGGCGCGCCGCSSGSGGCSSGGSGGSGSGCCGGCSSCGCCSGCCCCSSGCSCC 699

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | |     : ||:|  ||  | |  ::|| : | : : | |:|   :: ||::: |  |: 
Db        700 CCCGCGCGSCGGSCCGCGCSCSGGCSSCGCSCCCSCSCCCCSCCCCSSCGCSSSSGCGGS 759

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              | | :|  : |||:   : ||: |||||:||:::|||:|::|:  | : :|::   |
Db        760 GGCSSGCGSCGCGSCGGSGCGSCGGCGGSCGSSSGCGSGSSGSGCGSSSSGSSCGGGSGC 819

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              ::||::  : || :  :     |::|     ||  :: || : || :|  |  |:
Db        820 SSGCSSGGSSGCCSSGSGGCCCCSSGCGGCGCGSCSSCGCGSGGCCSGSCCGCCS 874


RESULT 2
CNS0091P
LOCUS       CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             26.6%;  Score 78.6;  DB 10;  Length 925;
  Best Local Similarity   16.3%;  Pred. No. 0.00011;
  Matches   39;  Conservative  137;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              | :|: |:|: :::::  : :: :: : :::: :::| ::::|:::::: :::::| :| 
Db        687 GGSGSTGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSSSTSSBBSCTSTS 746

Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCT 122
              :: ::: :::||:|| |  :::::::  ::::: ::|: : :::::::| :::::: : || 
Db        747 SSSSSSYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCS 806

Qy        123 CCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGC 182
              || : |   : :: :  |:|:|| :::::|   : :| ||  |:::|   :||::| :::
Db        807 CCCYMCTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTNGMBGTSSACSSS 866

Qy        183 CTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              : : : : :::::::: :::::::::|::|::: ::: : ::::| : :|: : :|  :| 
Db        867 SSSCSSSSVSSSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVSSGSGSGSGSVSG 925


RESULT 3
CNS006XK
LOCUS       CNS006XK                935 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14N09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             26.5%;  Score 78.2;  DB 10;  Length 935;
  Best Local Similarity   34.9%;  Pred. No. 0.00013;
  Matches  103;  Conservative   84;  Mismatches  107;  Indels    1;  Gaps    1;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              |||  |:  |   |:| |  |  : :  ::| ::|::| :||::::|: || : | ||: 
Db        617 GCGGCGSGCGCGCGSGGGCGGGCSSSGGSSGSSSGGSSGSSCGSSSSGSGGGGSGGGCGS 676

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                || | ||:|: ||:|||::||| || :|||:| | |: ||:|::| || :||:: ::
```

```
Db        677 GGCGGCGCGSCSGCGSCGCSSGCGSGCGSGCGSGGGCGSCGCSCSSC-GCGSGCSSGSSG 735

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | ||:|  : : :::  :  : |||| :: |:   |::|   |:| || ::    :| |
Db        736 CSCCSCCGSCSCSSSSGSCCSCGCGCGMSMCSGCGGSSCGGCCSGCCGCSSCGCCSGGCC 795

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              : |  |  | : :|| |   |:::| :|:| |: | |||:| : ||   :::|: ||  :
Db        796 SGCCGGCGCGSGSCGGCGCGGSSSGGSGSCCGSCGGGCGSCCSGGGCGCSSSCSGCGCGS 855

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              ||::|| :    :   :  :|  :::  :| :|  |:|:  |  : || :| :|:
Db        856 GGSSCCGSGGCGSGCGSGSSGGGSSSSGSGSSGGCGSGSGGGSCSGGGGSCCSCS 910


RESULT 4
CNS0091P/c
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match            26.1%;  Score 77;  DB 10;  Length 925;
  Best Local Similarity  14.3%;  Pred. No. 0.00021;
  Matches   39;  Conservative  153;  Mismatches   81;  Indels    0;  Gaps    0;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              ||: :::: :  :::: :   :  : ::: : ::: ::: |   |:| ::::::::::
Db        804 GASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSSA 745

Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCT 122
              :  ::::: : :::: ::::::  ::|::::|::: ::  : :::::::: :| :|:|||||
Db        744 SAGSVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSSASCASCSCCCT 685

Qy        123 CCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGC 182
```

```
              : :| |: : :::  :  : ::::|:::: :|::: : :: ::::: |:::: |:|  ::
Db        684 SWSCSCSTSASMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSSSGSSSSSSGSACGBS 625

Qy        183 CTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGG 242
              : :|| : ::: :  : ::::::: |:::|: || |:|||:|:::: : ::  : ::   ::
Db        624 MSSGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSCSS 565

Qy        243 GCCCACCTGCGCTTCCAGGACCGGACCGTCGTG 275
              ||:|:|: ::         :| : |::||::: |
Db        564 GCMCRCSCSSAAAAAASCVAASCGMMCGKSKSG 532


RESULT 5
CNS0052P/c
LOCUS       CNS0052P                 844 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056652
VERSION     AL056652.1  GI:4932342
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR11P16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match            26.0%;  Score 76.8;  DB 10;  Length 844;
  Best Local Similarity  18.5%;  Pred. No. 0.00023;
  Matches   45;  Conservative  135;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         12 CTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCCCGGC 71
              :: : |:::   : :::::::|:| ::: :::|::|::: :| |:|| : :|| :::: :
Db        639 YYCSCCBYBCNTSCSYYSYSBCSGCSSSSSSSCSSGSSBYSCCCBCCCBGYCGCYSBVCS 580

Qy         72 GTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCTCGAG 131
              :::: :::::|:: :  :| ||:|: :: :::|:::::|||  |:::| :|::: |::  :
Db        579 SBSSTBSSVSGSBSSTSSGTGCKCBSSSSSBBCSSBSSCCCYCSSYCTYCYSSCCBYCCB 520

Qy        132 GGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGGACAG 191
              :|||:|:: :: ::|:  :::::  ::| ::: :|| ::: | |:   :  || ::: | :
Db        519 SGCBTBSYBSCSSGBTGBSSBSGGKSGGKSSGKGGGBKSGGKGYCYCBTYCCSBSSBCCS 460
```

```
Qy        192 CGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCACCTG 251
              ::::  ::| :||::: :||:::: | :::: :|: : :: ::  |  |::: :: :  :
Db        459 SSSSSTSKCCSGGSSSTKCGSSBBGGGSSSKTSGSSSCBSGSSSTGSSGSSKKYSYSTSS 400

Qy        252 CGC 254
              :::
Db        399 SSS 397


RESULT 6
AG057427/c
LOCUS       AG057427               1755 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-044A05.F, genomic survey sequence.
ACCESSION   AG057427
VERSION     AG057427.1  GI:16594886
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1755)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1755
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-044A05.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             25.7%;  Score 75.8;  DB 10;  Length 1755;
  Best Local Similarity   50.9%;  Pred. No. 0.0003;
  Matches  140;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy          8 CGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCC 67
              |||   |||   || |   |   |       || |||||  |||||||||||||      |
Db        748 CGGCGCGCGGCGCGCGCGNCGNNGNNGNCGCGCGCGCGNNGGCGGCGCGCCGCGCNCGCG 689

Qy         68 CGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCT 127
              ||     ||  ||     |  |||||    | ||   |||| ||||  |||| | ||
Db        688 CGCGCGNGCGCCGNNNCCNCGCGCGCGGCGCGCGCCGCGCGGCGCGCGNGCCCCCCGGCG 629

Qy        128 CGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGG 187
              ||  | || |   |  ||||| ||  |||    ||   ||   ||| |  |||| | ||
Db        628 CGCCGCCGGCCGCCGCCGCCCGCCGGCCGCCGCGGCCCCCGGNGGCGNGNNGCGCGCCCG 569

Qy        188 ACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCA 247
                |  | |  |  || |||||  ||   | |  |  | ||   ||    ||  |   | |
```

```
Db        568 GCGNCNCCCCGGCGCGGGCGNCCGNGCGGGGCGGNCGGGCNNGGNTCCCGCCGCCCCTCG 509

Qy        248 CCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCG 282
              || ||||   | |   ||| |||      ||| ||
Db        508 CCCGCGCCCGCCGCCGCGGCCCGNGCGCCGGGGCG 474


RESULT 7
AG043044
LOCUS       AG043044               1134 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-021D20.R, genomic survey sequence.
ACCESSION   AG043044
VERSION     AG043044.1  GI:16571769
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1134)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1134
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-021D20.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             25.6%;  Score 75.4;  DB 10;  Length 1134;
  Best Local Similarity   52.4%;  Pred. No. 0.00038;
  Matches  151;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          8 CGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTCC 67
              ||    | ||  || | |  ||| ||| |    | | ||  | ||| | ||||  || ||
Db        486 CGCCGNGGGCNGCGCGCCCGCCCGGCCGCCNCGCCCCCCGCCCCGCCCCCCGCCCCGCCC 545

Qy         68 CGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGCT 127
              |  || | ||||| ||| || | |||||   | ||||  ||  || | || |    |||
Db        546 CCCCGCCNCCGCGCGCGCGCCGCCGCGCCCCGNGCGCCCCGCGCCCCCGCGCGGCNCGCC 605

Qy        128 CGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCGG 187
              | ||  |   |      || |||  |||||  |||| ||   |||  |  |||  ||
Db        606 NCCGCGCCCCCCCCGCGCGCCCGCGGCCGCGCCNGCGGCCGGCGGCGCCGCCCGCGCCGC 665

Qy        188 ACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGGGCCCA 247
              | || ||  | | |  || |   ||| |||   |    |  || | |   |  ||||
Db        666 GCCGCCCGGGCGCCGCCGCCGCGCGCGCCGCGCGCCGCCCGCGGCCCGNGCCGCCCC 725
```

```
Qy        248 CCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              | ||||  ||  |  ||  ||  || | ||  |  |   |  | |||
Db        726 GCCGCGCGCCCGCGCGCGCGCCCNCGCGCGGCGCCCGGNCGGCCCGCG 773


RESULT 8
AG063529/c
LOCUS       AG063529               1306 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-052D09.F, genomic survey sequence.
ACCESSION   AG063529
VERSION     AG063529.1  GI:16615331
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1306)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1306
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-052D09.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             25.4%;  Score 75;  DB 10;  Length 1306;
  Best Local Similarity   55.6%;  Pred. No. 0.00044;
  Matches  164;  Conservative    0;  Mismatches  130;  Indels    1;  Gaps    1;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGGCGGCGCGCCGC 60
              |||    |||   |||||     | ||  || ||| || | ||  ||  |  ||| ||
Db        733 GCGGCCCCGGCGCGCGCGCGCCCGCCCGGCGGCCGCGCGCGGGCCTGGCGGCGGCGGCGG 674

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                |  |   || || ||||| || |  ||||| |     ||||| ||  |  ||||  |
Db        673 CGGGGGCGCCCGGCGGCGCGGCCGCGGCGGCGCCCCGCCCGCGCGCCGCACGCCGGCAGC 614

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | |||| |     || ||  |  ||| |||||  |||| ||| |||| |||    | |
Db        613 CGCCGCGCCGCCCCGGCGCGCGCCGCGCGGCGCGGGCGCGCGCCGGCGCGCGACGCGCGG 554

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||| |   | |||||    ||  | |||||||  |||||  |   |  |  |||  ||
Db        553 GCCGCCCGC-GCGCGCGGCCGCCGCCGGGCGGGCGCGCGTCCGGGCCCCCGCCTGCGGGC 495

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              | ||||  ||||||   | || ||   |   || || |||      || |  |||

Db        494 GCGCCCGTCTGCGCGGTCGGGGCCCCGCGCGCGCGGCGGCTTGCGGGGCGGCGCG 440


RESULT 9
CG769390
LOCUS       CG769390               1444 bp    DNA     linear   GSS 29-OCT-2003
DEFINITION  TcB52.3_G08_SP6 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CG769390
VERSION     CG769390.1  GI:38022569
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 1444)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1444
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN

  Query Match             25.4%;  Score 75;  DB 10;  Length 1444;
  Best Local Similarity   55.9%;  Pred. No. 0.00044;
  Matches  123;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy          7 GCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTTCGTC 66
              |||| | ||||| ||  |  | |    |||  | | |||||||||  |||||||  || |
Db        347 GCGGCCGGCGCGGCGGCGCGCGCNCGGCGGGGGGNGGGCGCGGGCGCGGCGCCGCGGCGGC 406

Qy         67 CCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCTCCGC 126
                ||   |||| || ||| ||  |  |||| || |   ||  | |  ||   | | |
Db        407 GCGNGCGCGCCNCGCGCGCGCNCGGCCGCGGGGGGGCGGGGCGGCGNCGCGGCGCCCGCG 466

Qy        127 TCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGCCTCG 186
              | |||||  |  | ||  |  |        |  |||||    | |||  | ||| | ||
Db        467 CGGCGGGCGCNGCGCCGCCGCGCGGCNNGTGGNGCGCGGNGNCGNGCGCNGCGCGGCGCG 526

Qy        187 GACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGG 226
              |   | ||| |  || | ||| | |||||||| |   |||
Db        527 GCGCGNGCGNCGNCGCGCGCGCGGGGCGGCGCCGGCGCGG 566


RESULT 10
AG430101
LOCUS       AG430101               1552 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-304P19.T7, genomic survey
            sequence.
ACCESSION   AG430101
VERSION     AG430101.1  GI:48073164
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1552)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1552
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-304P19.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             25.4%;  Score 75;  DB 10;  Length 1552;
  Best Local Similarity   53.6%;  Pred. No. 0.00043;
  Matches  156;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||  |||  | | ||||| |  ||    |||  | |  | ||||||| | |    | |||
Db        895 GCCGGGGGCGGCGGCGCGGCCGGGGGGGCCGGGCCGGGGGGCGCGCCGCCCCGCCCCCGG 954

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                | |||||    |  || ||| |  | | |||||| || | |  || ||||  ||  ||
Db        955 GGGGGCCCGGGCGCCGCGGGGGGGGCCGGCCGCGCGGGGGGGGGCGCCGCCCCGGGGGCC 1014

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
               |||   | || || |   | | | || |||  |||||  ||| | | |||     | ||
Db       1015 GGCCGGCGGGGGCCGCCCGGCGGGGGCCCGCGGGCGCGCCGGCGCGGGGGCGGCGCGGGC 1074

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              |      | | | ||    | |||  |  |||||||  |  |||  |      ||||  |  |
Db       1075 GGGCGCGGCGGGGCCGGCGCGGCCGGCGGCGGCGCGGGCGCGGGGCGGGCGCCGCGGGCG 1134

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGC 291
               |||||    |||    | || | ||   |    |||| ||  |  || ||
Db       1135 CGGCCCGGGCGCGGGCGCGGGGCGGGGGGGGGCGGGGGCCCCCGGGGGCGC 1185


RESULT 11
CNS010EW
LOCUS       CNS010EW                1009 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03P19 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098882
VERSION     AL098882.1  GI:5610493
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03P19"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             25.4%;  Score 74.8;  DB 10;  Length 1009;
  Best Local Similarity   34.6%;  Pred. No. 0.0005;
  Matches  102;  Conservative   73;  Mismatches  120;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              |:| |::  ||  | : :  | || || | |  :: :|   | | : : | ||||:|| :
Db        620 GSGCGSSGCGTRGGGSRKSGGGGGGCCSCGGGGVSCSCCSCCCCCSCSCCCGCGCKCCMS 679

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
                |  |||  |   ||:|:| ::|  | || |:: | |: | | || ||:|||| :|:
Db        680 CCCCCCCCCCCCCGGCBGSGMSYGGCCGGGGGSSGGGGSGGKGGGGGGGSCCGCASCSGG 739

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
               |  |||  | :  ::     :   ::::    : | ::| |: | |||||||: :  :|
Db        740 CCGCGCSGGYSCCSSSGCSCSSCGSSSSCCGCSCCCSSCSCSNGCGCGTGCGCSSSCCSC 799

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
               || ||  |  || | |   ::: :|:: :::||:: ::||| ||     |:||  |  |
Db        800 CCCSCGCGCGCCGGGSCGCGSSSCSCSSCSSSCGSSCSSGCGTCGCCGCCGYCGNGGGSG 859

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGCAGCG 295
              :|: |  || :: :   |  : ::  ::|::|  :::| ||: :  :| :: |::
Db        860 SGSGCGCCCCSSCSCCGCGCKSYSCCMSCSKCNCSSSGCCCSSSGGSGGSSGGSS 914


RESULT 12
AG076820
LOCUS       AG076820                931 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-071C06.R, genomic survey sequence.
ACCESSION   AG076820
VERSION     AG076820.1  GI:16628622
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
```

```
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 931)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .931
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-071C06.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             25.3%;  Score 74.6;  DB 10;  Length 931;
  Best Local Similarity   55.9%;  Pred. No. 0.00055;
  Matches  162;  Conservative    0;  Mismatches  125;  Indels    3;  Gaps    1;

Qy          2 CGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCT 61
              || | || | |   | | || ||  |  |   | || |  ||  | | | || | ||||
Db        480 CGGGCGCCGGCGCGGGGGCGGGGCGCGGCCCGCNGCGGGCCGGCCCGCCCGCCCCCCGCG 539

Qy         62 TCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCC 121
               |   ||| |  ||||| | ||| |   |  ||| | || |||| |||  | || ||  ||
Db        540 CCCCGCCGCCCCCGCCGGGCGCGGGGCGCGGCGGGCGGCGCGCCGCGCGCGGCCGCGGCC 599

Qy        122 TCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCG 181
              |||  ||  || | ||    ||||||||||  |||| ||||||   ||| |  | |||
Db        600 GCCGGGCGGCGGGGGCGCCGGGCGCCCGGCGCCGGCGCGCGCGGG---GCGGGGCGCGCG 656

Qy        182 CCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGG 241
              |  |||   |||||  |   | |   |||||||| | ||||  | | ||    ||   |
Db        657 CGGCGGCGGGCGCGCGCGGCGCGCGCGGCGGCGGGCCCGCGCGCGGCGGGGGCGCGCGCG 716

Qy        242 GGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGGTGC 291
              |||  |   ||||  ||  |     |  || ||  |||      |  || ||
Db        717 GGCGGAGGGGCGCCGCCGCGCGGCGGGCGGCGCCGGGCGGCGGGGGGCGC 766


RESULT 13
CNS0072Q/c
LOCUS       CNS0072Q                 932 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066742
VERSION     AL066742.1  GI:4945205
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14B09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             25.2%;  Score 74.4;  DB 10;  Length 932;
  Best Local Similarity   35.6%;  Pred. No. 0.0006;
  Matches   99;  Conservative   69;  Mismatches  110;  Indels    0;  Gaps    0;

Qy          3 GAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGCTT 62
              | :::::    :|||:    |: |  | : |: |||| |  ||:||  |:|: |||:|
Db        917 GCSSSSSCGGGSCGCSSGCSGSCCGCCGSSCSCGCCGCGGCCGSGGCGGSCSGGCCSCGC 858

Qy         63 CGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCCCT 122
               :  :||       ::| || |: :: | ::| ::  : ||:|   ::|:::: ||:|::
Db        857 GSSGSCGCGCCGCSSGGGGCCSCSSGGCSSCCSSGCSSGCSCCCGSSCSSSSCGCSCSSC 798

Qy        123 CCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGCGC 182
              :||| :: |: || ||    :|  :||  :| |::|| : || ::: |||:: || : |
Db        797 SCGCSSSCGSCCGCCGSCGCSCCGSCGCGSCGCSSGCCSCCGCSSSSGCGSSGGGGSGGG 738

Qy        183 CTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATGGG 242
              :|   |||:: ||  :| | |: ||:||    || |:  :   |||:| ||  |
Db        737 GSGSGGGSGCGSSGCCSGSGCGCGGSCCGSCGCGCGGGGGSGCSSGGGGCSGGCGSSGCC 678

Qy        243 GCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGC 280
              | |: :| :| :  ||  :  |||  || |: | |:|:
Db        677 GGCSCSCCSCCSSGCCSCSSGCGGCGCGCCSCGCGSGS 640


RESULT 14
CNS006XK/c
LOCUS       CNS006XK                 935 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
```

```
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14N09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             25.2%;  Score 74.4;  DB 10;  Length 935;
  Best Local Similarity   34.6%;  Pred. No. 0.0006;
  Matches  101;  Conservative   83;  Mismatches  104;  Indels    4;  Gaps    1;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              | | |:|::: : |||:|  : :| :| ||: :|  :|: || ::|: ::::| |::|:|
Db        930 GGGGGSGSSSGSSGCGSGCGSGSGGSCCCCSGSCCCSCSCGCCSSCSCSSSSCCCSSCSC 871

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTG----CAGGCGCGCGAGGAGCGCGGCGGCCCGCGG 116
                :  ||: | : | :|||| : : :    | :| :||  |: | :|: |:::|||||
Db        870 GCSCGCCSCGGSSCCSCGCGCSGSSSGCGCCCSGGSCGCCCGSCGGSCSCCSSSCCGCGC 811

Qy        117 CCCCCTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTG 176
              | :|:  |||  | :|||   |  : ||| |:|||   : |||  ::: ||| |:| : :
Db        810 CGSCSCGCGCCGGCSGGCCSGGCGSSGCGGCSGGCCGSSCCGCSGKSKCGCGCGSGGSCS 751

Qy        177 GAGCGCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTC 236
              : : : :  |:   ||::  :  : | |::|:|||:|| | : ||:|  :   :: || |
Db        750 SSSGSGSCGGSGGSGCSSCSSGCSCGCGSSGSGCGSCGCCCSCGCSCGCSCGCSSGCGSC 691

Qy        237 GATGGGGCCCACCTGCGCTTCCAGGACCGGACCGTCGTGGGGGCCGAGCAGG 288
              |  |: || |  | :|||  :|    |::: | : |:  :::|::  : :|
Db        690 GCSGSCGCGCCGCCSCGCCCSCCCCSCSSSSCGSSCSSCCSSSCSSCCSSSG 639


RESULT 15
BQ939301
LOCUS       BQ939301               1185 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8893358 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6399496 5', mRNA sequence.
ACCESSION   BQ939301
VERSION     BQ939301.1  GI:22354779
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1185)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13900  row: k  column: 17
            High quality sequence stop: 326.
FEATURES             Location/Qualifiers
     source          1. .1185
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6399496"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Co24"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             25.1%;  Score 74;  DB 5;  Length 1185;
  Best Local Similarity   54.9%;  Pred. No. 0.00068;
  Matches  146;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy          1 GCGAGGGCGGTCTGCGCGTCGAGGTCCTCCGTCCGGCCGAGCGCGCGGGCGGCGCGCCGC 60
              ||   | |||   |||||    || |   || |||| || | | |  |  |   || ||
Db        428 GCCCCGCCGGGGGGCGCGCGCGGGCCGCGCGCCCGGGCGCGGGGGGCGCGGCGCCGGCGT 487

Qy         61 TTCGTCCCGGCGTCGCCGCGGGCGTGCAGGCGCGCGAGGAGCGCGGCGGCCCGCGGCCCC 120
               ||  | | ||   | ||  |||  |  |||| |   | |  ||||| |||   ||||
Db        488 CGCGCGCGGCCGCTCCGGCCCGCGGACGCGCGCCCCCCGGGGCCGGCGCCCCCGAGCCCG 547

Qy        121 CTCCGCTCGAGGGCGTCGATCTGCGCCCGGCGCACGCGCTCGCGGGCGTGCGCGTGGAGC 180
              | || |  | ||    |||   |||| || |   |  || |||| ||| || ||  | ||
Db        548 CCCCCCGGGCGGCGCGCGACGCGCGCGCGACCGCCCGGCGCGCGCGCGGGCCCGCCGCGC 607

Qy        181 GCCTCGGACAGCGCGTCCTCGGGGGCGGGCGGCGGCGCGGCGCCGGTCAGGCCGTCGATG 240
              ||  ||  ||| ||  ||| |||| || |||||||  | || | |  || ||| |
Db        608 CGGCCGCGCACGGCGGCCGTGGGCGCGGCGGGGGGCGCGGGCCGGGGCGGCGCGCCGAGG 667

Qy        241 GGGCCCACCTGCGCTTCCAGGACCGG 266
              | ||| |  |||     || ||||
Db        668 CCGGCCGCGCCCGCGCGTCGGTCCGG 693


Search completed: April 28, 2006, 12:17:54
Job time : 1464.06 secs
```